Problèmes d'identification combinatoire et puissances de graphes

David Auger

Thèse de doctorat

Problèmes d'identification combinatoire et puissances de graphes

David Auger

7 Juin 2010

Directeur de thèse : Olivier Hudry

Télécom ParisTech

Sommaire

0 Préliminaires **5**

I Identification des sommets dans les graphes **9**

1 Codes identifiants **11**

2 Structure des graphes sans jumeaux **24**

3 Complexité algorithmique et codes identifiants **30**

4 Systèmes de contrôle **36**

II Puissances de graphes **50**

5 Problèmes extrémaux **53**

6 Racines carrées de graphes **59**

III Annexes **65**

A Parameters in twin-free graphs **66**

**B Induced Paths
in Twin-Free Graphs** **84**

C Longs cycles dans les graphes sans $(1, \leq 2)$-jumeaux **90**

D Complexity Results for Identifying Codes in Planar Graphs **105**

E Minimal Identifying Codes in Trees and Planar Graphs **124**

F Watching systems in graphs **139**

G Maximum Size of a Minimum Watching System **155**

H Sizes of G, G^r, $G^r \setminus G$, Undirected Case **187**

I Sizes of G, G^r, $G^r \setminus G$, Directed Case **199**

J On the Square Roots of a Graph **213**

Introduction

Le premier objectif de recherche qui m'ait été proposé par mon futur directeur de thèse Olivier Hudry, qui encadrait alors mon stage de Master 2, a été de tenter d'étendre une propriété des graphes sans r-jumeaux. Ce résultat, que j'ai prouvé de façon assez complexe (une meilleure preuve a été obtenue par la suite, voir 2.10), se trouve déjà au carrefour des deux thématiques que nous avons développées dans le cadre de cette thèse :

– les problèmes d'identification des sommets dans les graphes, car les graphes sans r-jumeaux forment une classe de graphes admettant certains systèmes qui identifient leurs sommets, que nous appelons codes r-identifiants ;
– les problèmes de puissances de graphes, et plus généralement de distance dans les graphes, car le «r» de «r-identifiant» signifie «à distance r».

Au départ, notre étude des puissances de graphes avait pour but d'établir certaines propriétés liées au codes r-identifiants, puis nous avons développé cette thématique pour son intérêt propre. Par conséquent, cette thèse se compose de deux parties, la première regroupant les résultats liés à l'identification des sommets dans les graphes, et la seconde les résultats ayant trait aux puissances de graphes. La plupart de nos résultats ont été publiés dans des articles de recherche ou ont été soumis pour publication ; nous fournissons tous ces articles dans des versions préliminaires en annexe de cette thèse (annexes A à J), en anglais à une exception près.

Les six premiers chapitres exposent en français nos résultats regroupés par thématique, ainsi que l'arrière-plan nécessaire à leur compréhension. Bien souvent, nous ne fournissons pas les preuves de ces résultats après les énoncés mais renvoyons à l'annexe où on pourra les consulter. Nous avons ajouté quelques résultats n'ayant pas été publiés, ceux-là avec leur preuve, pour compléter notre étude.

Chapitre 0

Préliminaires

Sommaire

0.1 Graphes **5**
0.2 Chaînes et cycles **5**
0.3 Couplages **6**
0.4 Voisinages, distance, boules et sphères **6**
0.5 Diamètre et rayon **6**
0.6 Graphes partiels et sous-graphes **7**
0.7 Couverture et ensembles dominants **7**
0.8 Étoiles et graphes hyperoctaédraux **7**
0.9 Puissances de graphes **8**
0.10 Graphes orientés **8**

Nous donnons ici les définitions des objets qui seront utilisés et étudiés dans le reste de ce travail ; nous précisons aussi certaines notations. Nous ne pourrons redéfinir toutes les notions utilisées (*e.g.* les notions d'adjacence, d'incidence ou les différents niveaux de connexité dans un graphe) et nous renvoyons à [16] et [21] pour tout ce qui relève de la théorie des graphes.

0.1 Graphes

Par *graphe* on entendra en général un graphe non orienté fini, sans boucles ni arêtes multiples, c'est-à-dire un couple $G = (V, E)$ où V, l'ensemble des sommets de G, est un ensemble fini et E, l'ensemble des arêtes de G, est une partie de l'ensemble des paires d'éléments de V. Une arête $\{x, y\}$, où x et y sont des sommets de G, sera notée indifférement xy ou yx. Si G est un graphe, nous ferons respectivement référence à ses ensembles de sommets et d'arêtes par $V(G)$ et $E(G)$. Rappelons que l'*ordre* de G est le nombre de sommets de G alors que sa *taille* désigne son nombre d'arêtes. Nous renvoyons à [16] ou [21] pour la définition des graphes complets à n sommets ainsi que des graphes complets bipartis à $p+q$ sommets, que nous notons respectivement K_n et $K_{p,q}$. Dans le reste de ces préliminaires, G désigne un graphe non orienté.

0.2 Chaînes et cycles

Une *chaîne élémentaire* dans G est une suite de sommets distincts $v_1, v_2, \cdots v_k$ (où $k \geq 1$) de G tels que, pour tout $i \in \{1, 2, \cdots, k-1\}$, on ait $v_i v_{i+1} \in E(G)$. La longueur d'une telle chaîne est $k-1$ et nous la noterons plus simplement $v_1v_2\cdots v_k$. Les *extrémités* de cette chaîne sont v_1 et v_k ; on dira aussi que cette chaîne *joint*, ou *relie*, les sommets v_1 et v_k. Pour tout $k \geq 1$, on définit (à isomorphisme près) *la* chaîne à k sommets P_k (ou *la* chaîne de longueur $k-1$) comme le graphe à k sommets $v_1, v_2, \cdots, v_k$ dont les arêtes, au nombre de $k-1$, sont les v_iv_{i+1} pour $i \in \{1, 2, \cdots, k-1\}$.

Un *cycle élémentaire* dans G est une suite de sommets $v_1, v_2, \cdots v_k$, où $k \geq 3$, telle que $v_1v_2\cdots v_k$ soit une chaîne élémentaire et de plus $v_1v_k \in E(G)$. La longueur d'un tel cycle est k. On parlera *du* cycle $\mathcal{C}_k$ pour le graphe construit comme la chaîne P_k ci-dessus, auquel on ajoute l'arête v_1v_k. La *maille* de G est le minimum des longueurs des cycles de G.

0.3 Couplages

Un *couplage* dans G est un ensemble d'arêtes $M \subset E(G)$ qui sont deux à deux non adjacentes. Ce couplage est dit maximal s'il n'est pas strictement contenu dans un autre couplage de G.

0.4 Voisinages, distance, boules et sphères

L'ensemble des *voisins* d'un sommet x, *i.e.* l'ensemble des sommets adjacents à x dans G, sera appelé *voisinage ouvert* de x et noté $N_G(x)$. Le *voisinage fermé* de x, quant à lui, est l'ensemble $N_G[x]$ constitué des voisins de x ainsi que de x lui-même. Cette notation s'étend aux sous-ensembles A de $V(G)$ en posant

$$N_G[A] = \bigcup_{x \in A} N_G[x].$$

Étant donnés deux sommets x et y de G, la longueur minimale d'une chaîne reliant x à y dans G est appelée *distance de x à y dans* G et notée $d_G(x, y)$. Si une telle chaîne n'existe pas cette distance est considérée infinie. Si G est connexe, il est immédiat de constater que l'application d_G constitue une distance sur $V(G)$.

Pour tout sommet x et tout entier $r \geq 0$, on note $B_G(x, r)$ la boule de centre x et de rayon r dans G, c'est-à-dire l'ensemble des sommets y de G tels que $d_G(x, y) \leq r$. La boule $B_G(x, 1)$ est donc égale au voisinage fermé $N_G[x]$.

0.5 Diamètre et rayon

Le diamètre $\mathrm{diam}(G)$ d'un graphe connexe G est la quantité

$$\mathrm{diam}(G) = \max_{x,y \in V(G)} d_G(x, y).$$

On appelle également *diamètre de* G une chaîne d'extrémités x et y tels que $d_G(x, y) = \mathrm{diam}(G)$. Le rayon $\rho(G)$ du graphe G est lui ainsi défini :

$$\rho(G) = \min_{x \in V(G)} \max_{y \in V(G)} d_G(x, y).$$

Autrement dit, $\rho(G)$ est le plus petit entier r tel qu'il existe un sommet c de G vérifiant $B_G(c, r) = V(G)$. Un sommet c tel que $B(c, \rho(G)) = V(G)$ s'appelle un *centre* de G.

0.6 Graphes partiels et sous-graphes

Si G est un graphe, un *graphe partiel* de G est un graphe H tel que $V(H) = V(G)$ et $E(H) \subset E(G)$; on obtient donc les graphes partiels de G en supprimant uniquement des arêtes à partir de G. En revanche, un *sous-graphe* de G, ou *sous-graphe induit*, est un graphe H tel que $V(H) \subset V(G)$ et dont les arêtes sont précisément les arêtes de G ayant leurs deux extrémités dans $V(H)$. Si $A \subset V(G)$, on peut donc parler du sous-graphe de *Ginduit par l'ensemble de sommets A*, que nous noterons $G[A]$; dans le cas où $A = V(G) \setminus \{x\}$, nous utiliserons l'abréviation $G \setminus x$. En combinant les notions de sous-graphe et de graphe partiel, on obtient celle de *sous-graphe partiel* de G, qui sont les graphes dont les ensembles de sommets et d'arêtes sont des sous-ensembles de ceux de G.

0.7 Couverture et ensembles dominants

On dit qu'un sommet x d'un graphe G *couvre* (ou *domine*) un autre sommet y de G si $x \in N_G[y]$. Un *ensemble dominant* (ou *couvrant*) de G est un sous-ensemble $\mathscr{C}$ de $V(G)$ tel que tout sommet de G soit couvert par au moins un élément de $\mathscr{C}$, ce qui est équivalent à dire que $N_G[\mathscr{C}] = V(G)$. Le cardinal minimal d'un ensemble couvrant de G est appelé *nombre de domination* de G et noté $\gamma(G)$. Nous renvoyons aux ouvrages [57] et [56] pour un panorama des différentes questions liées aux ensembles dominants.

0.8 Étoiles et graphes hyperoctaédraux

Présentons deux autres types de graphes auxquels nous ferons plusieurs fois référence. Si V est un ensemble fini et $v \in V$, l'*étoile* sur V de centre v est le graphe dont V est l'ensemble des sommets et dont les arêtes sont les vx pour $x \in V \setminus \{v\}$. À isomorphisme près ce graphe ne dépend que du cardinal de V et si $n \geq 1$, on pourra parler en ce sens de *l'étoile à n sommets*, qui n'est autre que le graphe biparti complet $K_{1,n-1}$.

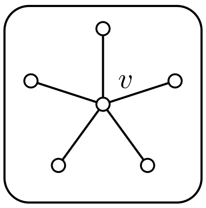


Figure 0.I – *Une étoile à 6 sommets, de centre v.*

Pour tout $n \geq 1$, le *graphe hyperoctaédral* d'ordre n, aussi appelé *cocktail party graph*, est défini (à isomorphisme près) comme étant un graphe complet d'ordre n auquel on a supprimé un couplage maximal. On note $\mathcal{H}_n$ ce graphe. Si $n = 2k$ est pair, on peut le décrire comme ayant pour ensemble de sommets

$$V = \{v_1, v_2, \cdots, v_k\} \cup \{v'_1, v'_2, \cdots, v'_k\}$$

et ayant toutes les arêtes possibles sauf entre les sommets v_i et v'_i pour $1 \leq i \leq k$.

Si $n = 2k + 1$ est impair avec $k \geq 1$, on obtient le graphe hyperoctaédral d'ordre n en ajoutant au graphe hyperoctaédral $\mathcal{H}_{2k}$ d'ordre $2k$ un nouveau sommet adjacent à tous les sommets de $\mathcal{H}_{2k}$ (voir la figure 0.II).

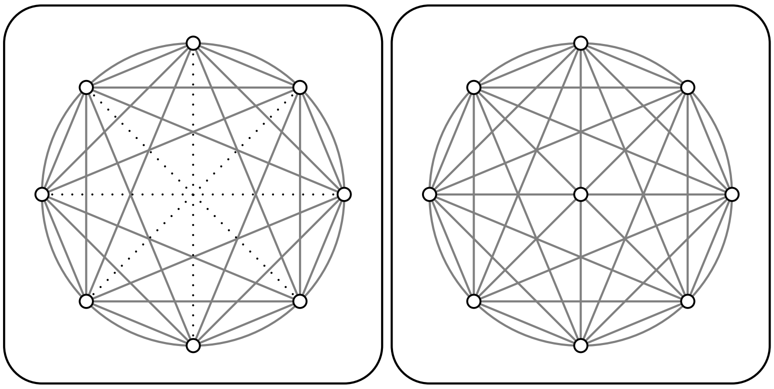


Figure 0.II – *À gauche, le graphe hyperoctaédral* $\mathcal{H}_8$ *d'ordre* 8 *: tous les sommets sont reliés deux à deux, à l'exception de quatre paires de sommets. À droite, le graphe* $\mathcal{H}_9$ *obtenu en ajoutant à* $\mathcal{H}_8$ *un sommet de degré* 8.

0.9 Puissances de graphes

Pour tout $r \geq 1$, la puissance r-ième d'un graphe G est le graphe noté G^r, ayant le même ensemble de sommets $V(G)$ que G et tel que deux sommets distincts x et y sont reliés par une arête dans G^r si et seulement si $d_G(x, y) \leq r$. Ainsi, G est un graphe partiel de G^2, qui est un graphe partiel de G^3, et ainsi de suite. Pour r supérieur ou égal au diamètre de G, le graphe G^r est complet. Si H est un graphe tel que $H^r = G$, on dit que H est une racine r-ième de G. Nous reviendrons en détail sur cette notion dans la partie II.

0.10 Graphes orientés

Nous n'envisageons des graphes orientés que dans la section 5.4 du chapitre 5. Un graphe orienté D sera pour nous la donnée d'un ensemble de sommets $V(D)$ et d'un ensemble $A(D)$ de couples de sommets, les *arcs* de D. Dans le cas présent nous supposerons que le graphe ne contient pas ne boucle (pas d'arc du type (x, x)) et pas d'arc multiples (au plus un arc (x, y) quand x, y sont deux sommets fixés) ; en revanche, D pourra contenir à la fois les arcs (x, y) et (y, x).

Un *chemin* de x à y dans D est l'analogue d'une chaîn de x à y dans un graphe non orienté : c'est une suite d'arcs

$$(x_0, x_1)(x_1, x_2) \cdots (x_{k-1} x_k)$$

où $x_0 = x$ et $x_k = y$, notée plus simplement $x_0 x_1 \cdots x_k$ et dont la *longueur* est k, son nombre d'arcs. La longueur du plus court chemin de x à y dans D est la *distance* $d_D(x, y)$; elle est *a priori* différente de $d_D(y, x)$ (ce n'est donc pas une distance au sens habituel). Les notions de diamètre, rayon et de puissance se définissent alors de façon analogue au cas non orienté. Rappelons enfin qu'un graphe orienté D est dit fortement connexe s'il existe, pour tout couple de sommets x et y de D, un chemin de x à y (et donc aussi un chemin de y à x) dans D.

Première partie

Identification des sommets dans les graphes

Les codes identifiants ont été introduits par M.G. Karpovsky, K. Chakrabarty et L.B. Levitin dans l'article séminal [65], afin de modéliser la détection et localisation de pannes dans les réseaux multi-processeurs. En un peu plus d'une dizaine d'années, ils ont donné lieu à de nombreuses publications (plus d'une centaine à ce jour) ; on pourra consulter une bibliographie quasi-exhaustive, méticuleusement maintenue à jour par A. Lobstein, à l'adresse électronique [73]. Cette liste recense aussi les articles concernant les codes localisateurs-dominateurs, une notion antérieure et assez proche (voir 1.5.1 à ce sujet). Citons également la thèse de J. Moncel ([74]) qui constitue une bonne introduction au sujet.

Les définitions concernant les codes identifiants et graphes sans jumeaux se trouvent dans le chapitre 1 ; nous ne pourrons y faire un état de l'art complet mais présenterons les résultats fondamentaux, ainsi que ceux directement liés à nos travaux. Les résultats de ce chapitre concernent essentiellement le cardinal minimal d'un code identifiant dans un graphe : nous y verrons quelques bornes inférieures et supérieures, ainsi que quelques graphes où ce cardinal minimal est connu exactement (ou presque). Les questions relatives aux graphes sans jumeaux, intrinsèquement liés aux codes identifiants, seront abordées dans le chapitre 2, puis le chapitre 3 traitera de la question de la complexité algorithmique du calcul de codes identifiants. Enfin, le chapitre 4 présente une nouvelle notion, celle des *systèmes de contrôle*, qui constitue une généralisation de la notion de code identifiant. Rappelons que la plupart des preuves de nos résultats ne seront pas données ici mais peuvent être consultées dans les annexes en troisième partie de cette thèse.

Chapitre 1

Codes identifiants

Sommaire

1.1 Introduction 11
1.2 Notations 15
1.3 Définitions 15
1.3.1 Identification des sommets 15
1.3.2 Identification des ensembles de sommets 16
1.4 Graphes sans jumeaux 16
1.5 Liens avec d'autres notions 18
1.5.1 Ensembles dominants et codes localisateurs-dominateurs 18
1.5.2 Puissances de graphes 18
1.6 Valeurs extrémales de $\gamma^{id}_{r,\ell}(G)$ 18
1.6.1 Borne inférieure 18
1.6.2 Borne supérieure 19
1.7 Chaînes et cycles 20
1.8 Grille du roi 20

1.1 Introduction

Les codes identifiants ont été introduits afin de modéliser des problèmes de localisation dans des réseaux, par exemple la recherche de pannes dans les systèmes de processeurs ([65]). Une autre application possible est la détection et localisation d'un incendie dans un bâtiment ([81]), que nous allons ici utiliser pour présenter le problème.

Prenons l'exemple d'un musée, représenté sur la figure 1.I (partie gauche), que nous voulons équiper de détecteurs de fumée. Ce musée est composé de huit pièces dénommées selon les huit premières lettres de l'alphabet ; certaines de ces pièces communiquent par une porte et d'autres non. Nous modélisons ce musée par un graphe G_1 comportant huit sommets correspondant chacun à une pièce ; deux sommets de ce graphe sont reliés par une arête si et seulement si les deux pièces correspondantes du musée communiquent par une porte. Ce graphe est également représenté sur la figure 1.I (partie droite).

Le cadre étant posé, supposons que nous disposions de détecteurs de fumée dont chaque pièce puisse être équipée. Ces détecteurs ont une portée d'une pièce, c'est-à-dire qu'ils détecteront de la fumée si un feu se déclare dans la pièce où ils se trouvent, ou dans une pièce immédiatement adjacente. Par exemple, un détecteur placé dans la pièce G pourra détecter si un feu se déclare dans les pièces E, F, G et H ; en revanche il ne détectera pas un feu en B ou en D. Le concept mathématique correspondant à celui d'un détecteur de

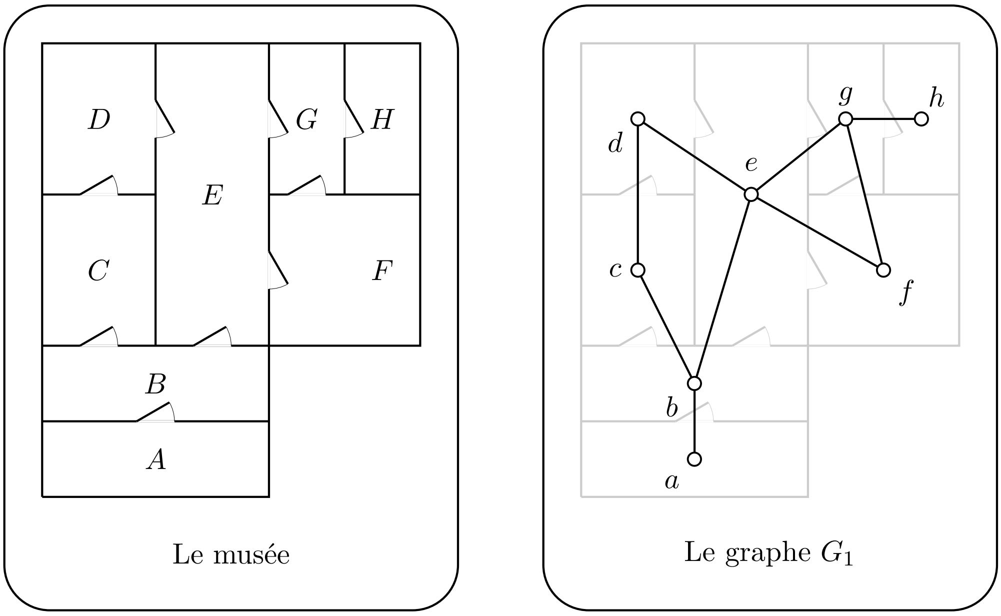


Figure 1.I – *Le musée où nous allons placer des détecteurs de fumée et le graphe G_1 qui lui est associé.*

fumée est celui de *mot de code*. Nous indiquons le choix d'un mot de code sur les figures en noircissant le sommet correspondant. Nous dirons ainsi par exemple que les sommets e, f, g et h sont *couverts* par le mot de code g : voir la figure 1.II.

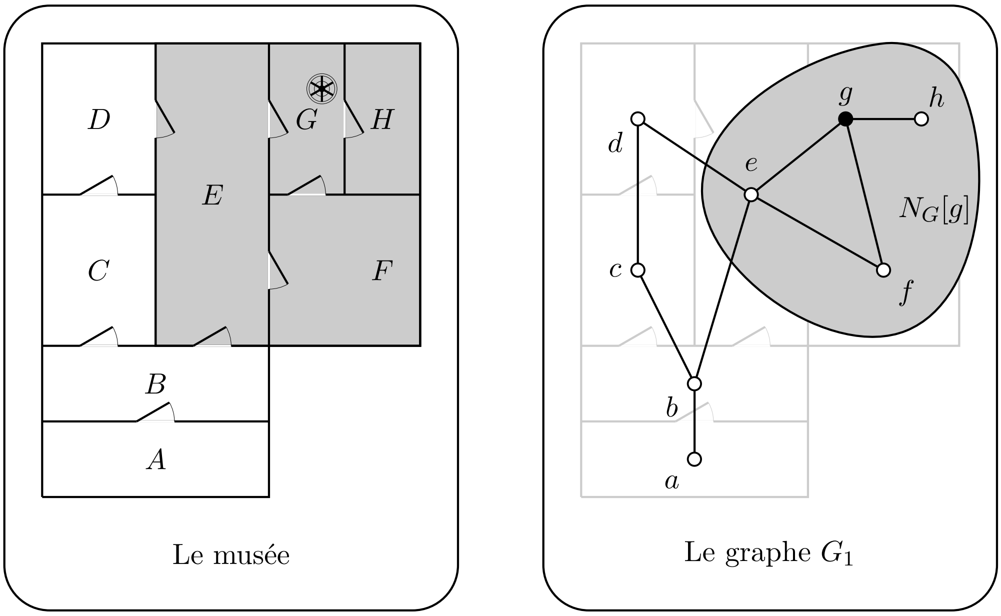


Figure 1.II – *Un détecteur de fumée dans la pièce G détecte les feux dans la zone grisée composée de G et des pièces adjacentes. Dans le modèle mathématique, le mot de code g couvre les sommets de son voisinage fermé $N_G[g]$ (zone grisée).*

Une première condition à remplir, si nous voulons pouvoir détecter un feu dans n'importe quelle pièce du musée, est de placer suffisamment de détecteurs de façon qu'une pièce, si elle ne comporte pas de détecteur, soit adjacente à au moins une pièce en comportant un. La notion correspondante dans le graphe G_1 est celle d'*ensemble dominant* (voir 0.7 page 7 pour une définition) : tous les sommets doivent être couverts par au moins un mot de code. Il est facile de se convaincre qu'ici trois mots de code au moins sont nécessaires pour couvrir tout le graphe. Une solution est présentée sur la figure 1.III ; il y en a d'autres.

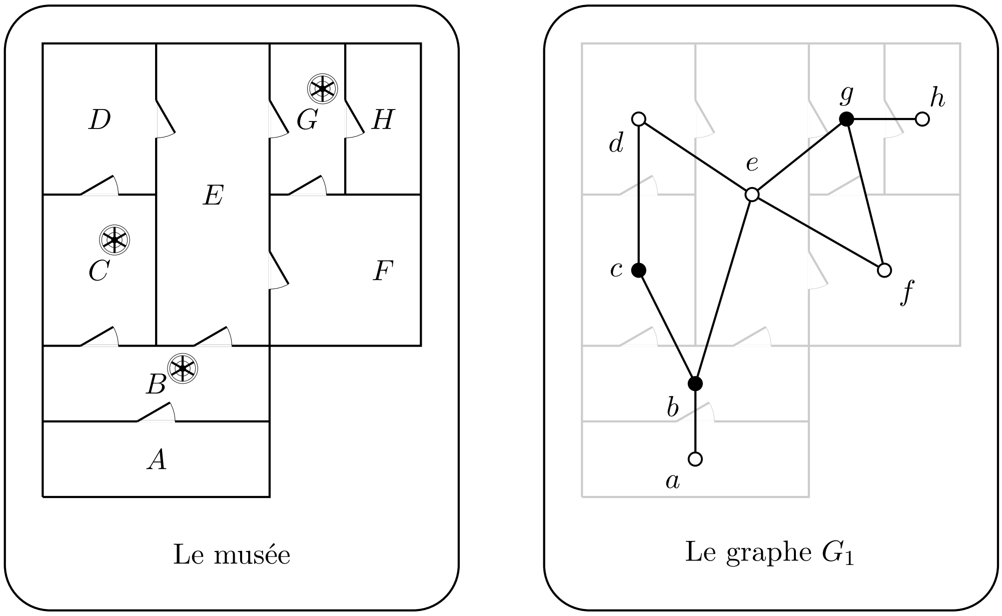


Figure 1.III – *Un exemple d'ensemble dominant composé des sommets b, c et g dans le graphe G_1. Dans le musée, les trois détecteurs de fumée correspondants suffisent à surveiller toutes les pièces.*

Déterminer un ensemble dominant de cardinal minimum dans un graphe donné est un problème difficile (voir le paragraphe 1.5.1). Ici, nous voulons cependant faire plus que surveiller l'intégralité du musée : nous voulons être capables de localiser les feux. Pour cela, supposons qu'au poste de commandement du gardien de musée se trouvent des voyants lumineux, un par détecteur de fumée, ces voyants s'allumant si le détecteur correspondant détecte de la fumée. En reprenant l'exemple de la figure 1.III, et en supposant qu'il y ait au plus un départ d'incendie dans notre musée, dressons une table indiquant, pour chaque pièce où un feu pourrait se déclarer, quels voyants s'allumeraient dans ce cas (table 1.I).

Comme on peut le constater, les mots de codes choisis forment bien un ensemble dominant car toutes les lignes de la table 1.I signalent au moins un voyant allumé. Considérons maintenant plusieurs hypothèses, toujours dans le cas où il n'y aurait qu'un seul départ d'incendie :

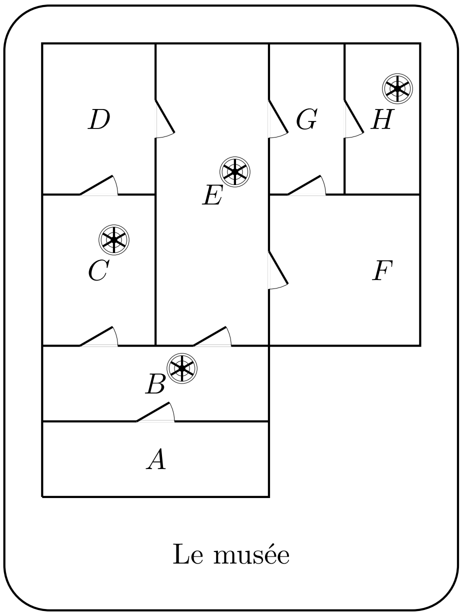
– supposons que seul le voyant B s'allume : c'est qu'alors un feu s'est déclaré dans le musée, et ce feu se trouve nécessairement dans la pièce A, car c'est l'unique cas où seul le voyant B s'allume ;
– de même, si seul le voyant C s'allumait, ou si les deux voyants B et G s'allumaient simultanément, alors nous pourrions localiser l'incendie (respectivement en D et en E) ;
– toutefois, dans les autres cas, par exemple si les voyants B et C s'allumaient, nous saurions qu'il y a un feu en B ou C, mais serions incapables de distinguer les deux cas avec ces seules informations, car les lignes correspondantes de la table 1.I sont

Pièce \ Voyants	B	C	G
A	•		
B	•	•	
C	•	•	
D		•	
E	•		•
F			•
G			•
H			•

Table 1.I – *Voyants allumés suivant les différents départs d'incendie dans l'exemple de la figure 1.III. On constate qu'aucune ligne n'est vide : le code forme un ensemble dominant.*

égales.

Dans le graphe G_1, nous dirons que le code choisi, à savoir $\mathscr{C} = \{b, c, g\}$, n'est pas suffisant pour *séparer* le sommet b du sommet c. Le même problème se présente avec les trois sommets f, g et h. Il nous faut pour ce faire davantage de mots de code, réalisant ainsi ce que nous appellerons un *code identifiant*, c'est-à-dire un choix de mots de code tel que la table correspondante, en plus de ne pas comporter de lignes vides, ne contienne pas deux lignes égales. Pour ce graphe, il nous faut au minimum quatre mots de code afin de réaliser un code identifiant. Un exemple optimal est donné sur la figure 1.IV, et on pourra vérifier les conditions requises pour être un code identifiant sur la table 1.II.



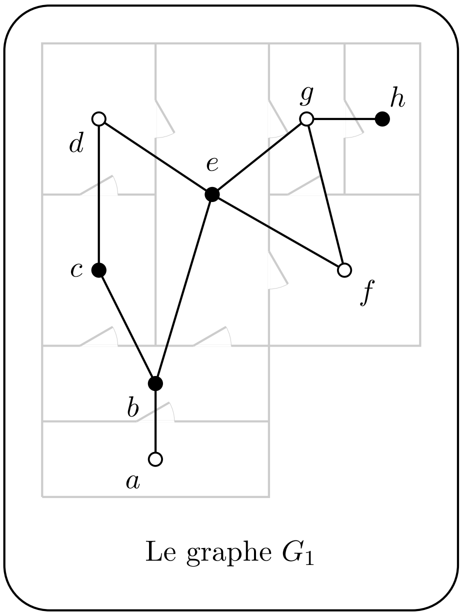


Figure 1.IV – *Un exemple de code identifiant composé des sommets b, c, e et h dans le graphe G_1 ; les mots de code sont les sommets noirs. Dans le musée, les quatre détecteurs de fumée correspondants suffisent à surveiller toutes les pièces et localiser les départs d'incendie.*

Nous généraliserons la notion de code identifiant dans un graphe à celle de code $(r, \leq \ell)$-identifiant où r et ℓ sont des entiers strictement positifs : intuitivement, un tel code remplit les mêmes fonctions qu'un code identifiant, à ceci près que la zone de couverture des mots

Pièce \ Voyants	B	C	E	H
A	•			
B	•	•	•	
C	•	•		
D		•	•	
E	•		•	
F			•	
G			•	•
H				•

Table 1.II – *Voyants allumés suivant les différents départs d'incendie dans l'exemple de la figure 1.IV. On constate qu'aucune ligne n'est vide et qu'elles sont deux à deux distinctes : le code est identifiant.*

de codes (*i.e.* la portée des détecteurs en nombre portes à franchir) est r et non plus 1, et que le code permet de détecter et localiser jusqu'à ℓ feux simultanés.

1.2 Notations

Comme nous venons de le mentionner, la notion de code identifiant se décline de plusieurs manières et on parlera de codes identifiants, de codes r-identifiants ou de codes $(r, \leq \ell)$-identifiants. Il se trouve que le terme «code identifiant» désigne aussi bien ce genre de codes dans son ensemble qu'un code $(1, \leq 1)$-identifiant ; on peut également parler de code 1-identifiant, car le terme «code r-identifiant» signifie la même chose que «code $(r, \leq 1)$-identifiant». Afin d'y voir clair dès maintenant et pour éviter de nous répéter par la suite, lorsqu'un paramètre lié aux codes identifiants (par exemple $\gamma^{id}_{r,\ell}(G)$, *cf.* 1.3.2) dépend d'un couple noté $(r, \leq \ell)$ ou (r, ℓ), nous le déclinerons (dans l'exemple en $\gamma^{id}_r(G)$ et $\gamma^{id}(G)$) avec la convention suivante :

Convention 1.1. *Lorsque $\ell = 1$ le couple $(r, \leq \ell)$ (ou (r, ℓ)) peut juste être noté r et lorsqu'on a également $r = 1$, la référence à $(r, \leq \ell)$ peut être omise.*

Ainsi «code $(r, \leq 1)$-identifiant» et «code r-identifiant» ont la même signification ; de même pour «code $(1, \leq 1)$-identifiant», «code 1-identifiant» et «code identifiant». Nous utiliserons la même convention pour les graphes sans $(r, \leq \ell)$-jumeaux ainsi que pour les systèmes de $(r, \leq \ell)$-contrôle au chapitre 4. Dans tout le chapitre, r et ℓ désignent des entiers naturels strictement positifs.

1.3 Définitions

1.3.1 Identification des sommets

Soit G un graphe non orienté. Un *code* de G est simplement une partie $\mathscr{C}$ de l'ensemble $V(G)$ des sommets de G, et ses éléments sont appelés *mots de code*. Soient x et y des sommets de G et $c \in \mathscr{C}$ un mot de code. On dit que :
- le sommet x est r-*couvert* par le mot de code c si $d_G(c, x) \leq r$;
- les sommets x et y sont r-*séparés* par le mot de code c si l'un de ces deux sommets est r-couvert par c et l'autre ne l'est pas.

On dira que x est r-couvert par le code $\mathscr{C}$ s'il est couvert par au moins un mot de code appartenant à $\mathscr{C}$, et nous parlerons de même de r-séparation de deux sommets par $\mathscr{C}$. Un sommet à la fois r-couvert et r-séparé de tous les sommets $y \in V(G) \setminus \{x\}$ par $\mathscr{C}$ sera dit *r-identifié*. On notera $I^r_{\mathscr{C}}(x)$ l'*ensemble identifiant de x*, qui est constitué des mots de codes qui r-couvrent le sommet x, c'est-à-dire

$$I^r_{\mathscr{C}}(x) = B_G(x, r) \cap \mathscr{C}.$$

Définition 1.2. *Un code $\mathscr{C} \subset V(G)$ est dit r-identifiant pour le graphe G si l'une des deux conditions équivalentes qui suivent est vérifiée :*

(i) tous les sommets de G sont r-identifiés par $\mathscr{C}$;

(ii) les ensembles $I^r_{\mathscr{C}}(x)$ pour $x \in V(G)$ sont tous non vides et deux à deux distincts.

Conformément à la convention 1.1, lorsque $r = 1$ on écrit *code identifiant* pour «code 1-identifiant», «$I_{\mathscr{C}}(x)$» pour «$I^1_{\mathscr{C}}(x)$», «séparés» pour «1-séparés», etc.

1.3.2 Identification des ensembles de sommets

Si $\mathscr{C}$ est un code du graphe G et $A \subset V(G)$ est un ensemble de sommets de G, on définit $I^r_{\mathscr{C}}(A)$ comme l'ensemble des mots de code r-couvrant au moins un sommet de A, c'est-à-dire :

$$I^r_{\mathscr{C}}(A) = \bigcup_{x \in A} I^r_{\mathscr{C}}(x).$$

On a donc, pour tout sommet x, la relation $I^r_{\mathscr{C}}(\{x\}) = I^r_{\mathscr{C}}(x)$, ainsi que $I^r_{\mathscr{C}}(\emptyset) = \emptyset$. Les codes identifiant les ensembles de sommets, ou codes $(r, \leq \ell)$-identifiants sont alors ainsi définis :

Définition 1.3. *Étant donnés deux entiers strictement positifs r et ℓ, un code $\mathscr{C} \subset V(G)$ est dit $(r, \leq \ell)$-identifiant pour le graphe G si les ensembles $I^r_{\mathscr{C}}(A)$ sont deux à deux distincts pour toutes les parties A de $V(G)$ de cardinal au plus ℓ.*

Remarquons que si $A \neq \emptyset$ on aura alors toujours $I^r_{\mathscr{C}}(A) \neq \emptyset$; pour $\ell = 1$, on retrouve donc la notion de code r-identifiant. Nous noterons $\gamma^{id}_{r,\ell}(G)$ le cardinal minimum d'un code $(r, \leq \ell)$-identifiant dans un graphe G, en appliquant la convention 1.1 qui donnera les abréviations $\gamma^{id}_r(G)$ et $\gamma^{id}(G)$.

1.4 Graphes sans jumeaux

Si deux sommets distincts d'un graphe G ont le même voisinage fermé, alors aucun mot de code ne pourra les séparer ; deux tels sommets d'un graphe seront appelés 1-*jumeaux*, ou plus simplement *jumeaux* (voir fig. 1.V). Ainsi, un graphe contenant deux jumeaux ne saurait admettre de code identifiant. Plus généralement, deux r-jumeaux dans un graphe G, où $r \geq 1$, sont deux sommets distincts x et y de G tels que

$$B_G(x, r) = B_G(y, r)$$

(voir la figure 1.VI pour un exemple de 2-jumeaux dans un graphe). Un graphe ne contenant pas de r-jumeaux est tout bonnement appelé *graphe sans r-jumeaux*.
Le lien précis avec les codes identifiants est donné par le théorème suivant (dont la preuve est élémentaire) :

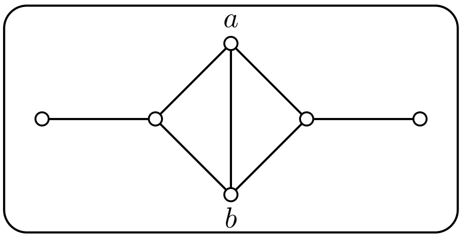


Figure 1.V – *Les sommets a et b sont jumeaux ; ce graphe ne peut admettre de code identifiant.*

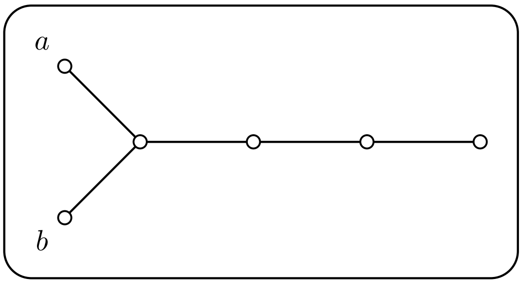


Figure 1.VI – *Un graphe sans jumeaux, mais où les sommets a et b sont 2-jumeaux.*

Théorème 1.4. *Un graphe admet des codes r-identifiants si et seulement s'il est sans r-jumeaux.*

Autrement dit, un graphe G admet des codes identifiants si et seulement si $V(G)$ est un code identifiant de G ; la figure 1.VII montre un exemple de code 2-identifiant dans un graphe sans 2-jumeaux. Le fait qui précède se généralise aux codes $(r, \leq \ell)$-identifiants :

Théorème 1.5. *Pour tous $r \geq 1$ et $\ell \geq 1$, un graphe G admet des codes $(r, \leq \ell)$-identifiants si et seulement si pour tous les ensembles distincts $A, B \subset V(G)$ de taille au plus ℓ on a*

$$\bigcup_{x \in A} B_G(x, r) \neq \bigcup_{x \in B} B_G(x, r).$$

Les graphes satisfaisant la condition du théorème précédent sont appelés *graphes sans $(r, \leq \ell)$-jumeaux* ; les jumeaux en question sont ici des ensembles de sommets que le code ne peut distinguer. Le chapitre 2 sera entièrement consacré à l'étude des graphes sans $(r, \leq \ell)$-jumeaux ; dans la suite du chapitre présent nous étudierons les propriétés des codes identifiants dans ces graphes.

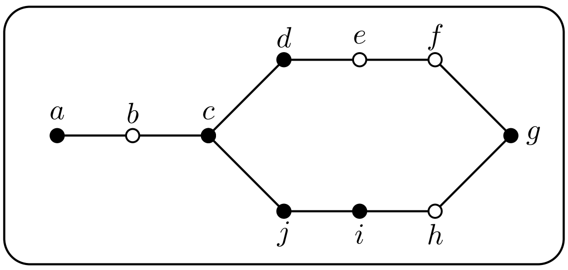


Figure 1.VII – *Un graphe sans 2-jumeaux muni d'un code 2-identifiant (sommets noirs).*

1.5 Liens avec d'autres notions

1.5.1 Ensembles dominants et codes localisateurs-dominateurs

Rappelons qu'un *ensemble dominant* dans un graphe G est un code $\mathscr{C} \subset V(G)$ tel que tout sommet de G soit couvert par au moins un mot de code (voir 0.7) ; un code identifiant est donc en particulier un ensemble dominant. Un code $\mathscr{C}$ dans un graphe G est dit *localisateur-dominateur* si les ensembles $I_{\mathscr{C}}(x)$ sont tous non vides et deux à deux distincts pour tous les $x \in V(G) \setminus \mathscr{C}$. On peut donc imaginer un code localisateur-dominateur comme un code identifiant où les mots de code seraient automatiquement identifiés. On remarque donc qu'un code identifiant est un code localisateur-dominateur, et que les codes localisateurs-dominateurs sont aussi des ensembles dominants. En notant $\gamma^L(G)$ la taille minimale d'un code localisateur-dominateur dans G et $\gamma(G)$ le nombre de domination de G, il vient donc

$$\gamma(G) \leq \gamma^L(G) \leq \gamma^{id}(G).$$

Les codes localisateurs-dominateurs ont été introduits par P.J. Slater en 1988 ([87]) et ont donné lieu à de nombreux développements. Nous renvoyons à [73] pour une bibliographie quasi-exhaustive les concernant, ainsi qu'au chapitre qui s'y rapporte dans [57].

Notons que calculer le cardinal minimum d'un ensemble dominant, ou d'un code localisateur-dominateur, dans un graphe donné constitue un problème $\mathcal{NP}$-difficile (voir respectivement [64] et [41]) ; nous verrons au chapitre 3 que c'est aussi le cas pour la recherche de codes identifiants.

1.5.2 Puissances de graphes

La notion de puissance de graphe (voir 0.9) est commode pour ramener l'étude des codes $(r, \leq \ell)$-identifiants à celle des codes $(1, \leq \ell)$-identifiants, puisqu'on a le résultat suivant (dont la preuve est immédiate) :

Théorème 1.6. *Pour $r \geq 1$ et $\ell \geq 1$, un graphe G est sans $(r, \leq \ell)$-jumeaux si et seulement si G^r est sans $(1, \leq \ell)$-jumeaux, et un code $\mathscr{C} \subset V(G)$ est $(r, \leq \ell)$-identifiant pour G si et seulement s'il est $(1, \leq \ell)$-identifiant pour G^r.*

Nous étudions cette notion plus en détail dans la seconde partie de cette thèse.

1.6 Valeurs extrémales de $\gamma^{id}_{r,\ell}(G)$

1.6.1 Borne inférieure

Une borne élémentaire, en fonction du nombre de sommets, est la suivante.

Théorème 1.7 (Karpovsky, Chakrabarty, Levitin [65])**.** *Soit G un graphe sans r-jumeaux d'ordre n. Alors*

$$\lceil \log_2(n+1) \rceil \leq \gamma^{id}_r(G).$$

S. Gravier et J. Moncel ([76]) ont caractérisé les graphes pour lesquels cette borne est atteinte, dans le cas $r = 1$. I. Charon, O. Hudry et A. Lobstein ont montré ([37]) que pour tout k satisfaisant

$$\lceil \log_2(n+1) \rceil \leq k \leq n-1$$

il existait un graphe sans jumeaux G d'ordre n tel que $\gamma^{id}_1(G) = k$.

1.6.2 Borne supérieure

Étant donné un graphe sans r-jumeaux G, quelle est la valeur maximale de $\gamma^{id}_{r,\ell}(G)$ en fonction de l'ordre n de G ? Le résultat suivant a été trouvé indépendamment plusieurs fois :

Théorème 1.8 (Gravier, Moncel [54], Bertrand, Charon, Hudry, Lobstein [33])**.** *Soit $r \geq 1$ et G un graphe sans r-jumeaux connexe d'ordre $n \geq 3$. Alors*

$$\gamma^{id}_r(G) \leq n - 1.$$

Le théorème 1.8 peut aussi être vu comme une conséquence du théorème suivant dû à A. Bondy, dont nous fournissons une esquisse de preuve.

Théorème 1.9 (A. Bondy [20])**.** *Soient*

$$S_1, S_2, \cdots, S_n$$

n parties distinctes d'un ensemble fini non vide X, avec $n \leq |X|$. Alors il existe $x \in X$ tel que les parties

$$S_1 \setminus \{x\}, S_2 \setminus \{x\}, \cdots, S_n \setminus \{x\}$$

soient encore distinctes.

Preuve. Par l'absurde : supposons qu'un tel élément x n'existe pas. Nous construisons un graphe non orienté dont les sommets sont les ensembles S_i, et nous ajoutons pour chaque $y \in X$ une arête dans le graphe entre deux parties S_i et S_j tels que $S_i \setminus \{y\} = S_j \setminus \{y\}$; deux telles parties doivent exister par hypothèse. Il est facile de constater qu'un tel graphe ne peut avoir de cycle, et doit donc avoir moins d'arêtes que de sommets : soit

$$|X| < n,$$

une contradiction. □

Ce théorème est optimal, comme le montrent sur l'ensemble $X = \{1, 2, \cdots, n\}$ les familles

$$\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}, \cdots, \{1, 2, \ldots, n\}$$

et

$$\emptyset, \{1\}, \{2\}, \{3\}, \cdots, \{n\}$$

qui comptent chacune $n + 1$ éléments.

Preuve du théorème 1.8. Les boules $B_G(v, r)$ pour $v \in V(G)$ sont deux à deux distinctes et de cardinal supérieur ou égal à 2 puisque G est sans r-jumeaux, connexe et compte au moins trois sommets. Appliquons le théorème 1.9 à la famille

$$\{B_G(v, r) : v \in V(G)\}$$

de n sous-ensembles de $V(G)$: il existe un sommet x tel que les n ensembles composant la famille

$$\{B_G(v, r) \setminus \{x\} : v \in V(G)\}$$

soient deux à deux distincts. Étant également non vides, on en déduit que $V(G) \setminus \{x\}$ est un code identifiant de G. □

Notons que ce théorème n'est valable que pour $\ell = 1$; lorsque $\ell \geq 2$, on peut avoir $\gamma^{id}_{r,\ell}(G) = n$, comme le montre l'exemple du cycle : pour $n \geq 4r + 3$ et $\ell = 2$ on établit facilement

$$\gamma^{id}_{r,2}(\mathcal{C}_n) = n.$$

1.7 Chaînes et cycles

Les graphes les plus simples pour lesquels on peut tenter de calculer les valeurs de $\gamma_{r,\ell}^{id}$ sont les chaînes et les cycles, et ceci constitue déjà un problème combinatoire assez complexe. La chaîne $\mathcal{P}_n$ est sans r-jumeaux dès que $n \geq 2r+1$. Des résultats partiels ont été obtenus dans [17], et le cas $r = 2$ a été traité dans [82]. Finalement Junnila et Laihonen ont traité les cas restants ([63]). Les différentes valeurs de $\gamma_r^{id}(\mathcal{P}_n)$ sont données par la table 1.III.

n	condition	$\gamma_r^{id}(\mathcal{P}_n)$
$n = 2r+1$		$2r$
$2r+2 \leq n \leq 4r+1$		$2r+1$
$n = 4r+2$		$2r+2$
$n \geq 4r+3$ avec $n = \frac{q(2r+1)}{2} + p$ où $1 \leq p \leq 2r+1$	q pair et $1 \leq p \leq r+1$	$\frac{q(2r+1)}{2} + p$
	q pair et $2r+1 < p$	$\frac{q(2r+1)}{2} + p - 1$
	q impair et $1 \leq p \leq 2r$	$\frac{(q+1)(2r+1)}{2}$
	q impair et $2r < p$	$\frac{(q+1)(2r+1)}{2} + 1$

Table 1.III – *Taille minimum d'un code r-identifiant dans $\mathcal{P}_n$*

La chaîne $\mathcal{P}_n$ ayant des sommets de degré 1, elle n'admet aucun code $(r, \leq \ell)$-identifiant dès que $\ell \geq 2$ (voir 2.4.1), sauf dans le cas élémentaire $n = 1$.

Passons maintenant au cas des cycles $\mathcal{C}_n$. Le cas $r = 1$ a été traité dans [55], et le cas $r = 2$ dans [82]. Le cas général a été terminé il y a peu. Les résultats sont résumés par les tables 1.IV et 1.V. Les références sont données en fonction du premier article ayant établi les valeurs.

n	$\gamma_r^{id}(\mathcal{C}_n)$	référence
$n = 2r+2$	$n-1$	[17]
$n \geq 2r+4$, n pair	$\frac{n}{2}$	[17]

Table 1.IV – *Taille minimum d'un code r-identifiant dans $\mathcal{C}_n$ pour n pair*

Le cycle $\mathcal{C}_n$ est sans $(r, \leq 2)$-jumeaux dès que $n \geq 4r+3$, son unique code $(r \leq 2)$-identifiant contenant tous les sommets. Enfin, dès que $\ell \geq 3$, le cycle $\mathcal{C}_n$ n'admet plus de codes $(r, \leq \ell)$-identifiants.

1.8 Grille du roi

La *grille du roi* est un graphe non orienté infini, dont l'ensemble des sommets est $\mathbb{Z}^2$, et tel que deux sommets distincts (i,j) et (i',j') sont adjacents si et seulement si $|i-i'| \leq 1$ et $|j-j'| \leq 1$; ainsi deux sommets sont adjacents si, sur l'échiquier infini $\mathbb{Z}^2$, un roi peut se déplacer en un coup d'un sommet à l'autre. Le figure 1.VIII représente un morceau de ce graphe. Nous présentons ici quelques résultats sur les codes identifiants de ce graphe,

n	conditions	$\gamma_r^{id}(\mathcal{C}_n)$	référence
$n = 2r+3$		$\left\lceil \frac{2n}{3} \right\rceil$	[55]
$2r+5 \leq n \leq 3r+1$	$pgcd(2r+1, n) = 1$ et $n = 2mp+1$ ou $n = (2m+2)p-1$ avec $n = 2r+1+p$ et $m \geq 2$	$\frac{n+1}{2}+1$	[63]
	$pgcd(2r+1, n) = 1$ autres cas	$\frac{n+1}{2}$	[63]
	$pgcd(2r+1, n) > 1$	$pgcd(2r+1, n) \left\lceil \frac{2}{pgcd(2r+1,n)} \right\rceil$	[63]
$3r+2 \leq n$	$pgcd(2r+1, n) = 1$ et $n = 2m(2r+1)+1$ ou $n = (2m+1)(2r+1)+2r$ avec $m \geq 1$	$\frac{n+1}{2}+1$	[55] et [89]
	$pgcd(2r+1, n) = 1$ autres cas	$\frac{n+1}{2}$	[55] et [89]
	$pgcd(2r+1, n) > 1$	$pgcd(2r+1, n) \left\lceil \frac{2}{pgcd(2r+1,n)} \right\rceil$	[55]

Table 1.V – *Taille minimum d'un code r-identifiant dans $\mathcal{C}_n$ pour n impair*

essentiellement pour les comparer aux résultats analogues pour les systèmes de contrôle (chapitre 4).

On définit les codes identifiants dans la grille du roi de manière analogue au cas fini : un code identifiant est un ensemble $\mathscr{C}$ de sommets de la grille du roi qui couvre tous les sommets et sépare toutes les paires de sommets distincts. On généralise de même au cas infini la notion de code $(r, \leq \ell)$-identifiant pour $r \geq 1$ et $\ell \geq 1$. Soit pour $n \geq 0$ l'ensemble T_n défini par

$$T_n = \{(i,j) : -n \leq i \leq n \text{ et } -n \leq j \leq n\}.$$

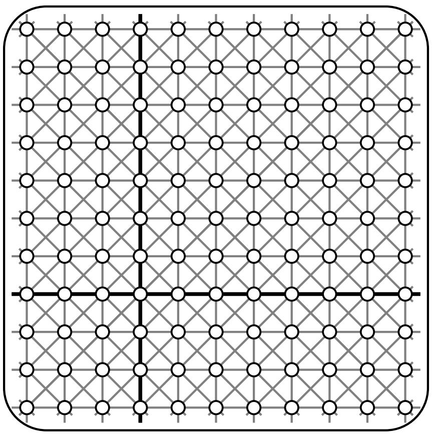


Figure 1.VIII – *La grille du roi (extrait).*

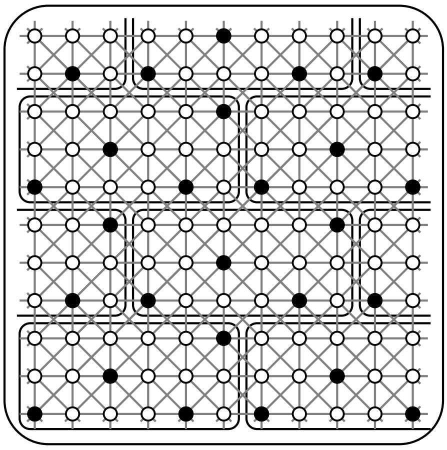


Figure 1.IX – *Un code 1-identifiant de densité minimale dans la grille du roi ; les mots de code sont les sommets noirs.*

La *densité* d'un code $\mathscr{C}$ de la grille du roi est alors définie par

$$\limsup_{n\to+\infty}\frac{|\mathscr{C}\cap T_n|}{|T_n|}=\limsup_{n\to+\infty}\frac{|\mathscr{C}\cap T_n|}{(2n+1)^2}.$$

Nous avons le résultat suivant :

Théorème 1.10 (Charon, Honkala, Hudry, Lobstein [32] & [40])**.** *Soit $D_{1,1}$ la borne inférieure des densités des codes identifiants dans la grille du roi. On a*

$$D_{1,1}=\frac{2}{9}.$$

La figure 1.IX représente un code 1-identifiant de densité optimale dans la grille du roi : comme la plupart des codes trouvés en pratique ce code est périodique.
Pour $r>1$, nous avons le résultat suivant :

Théorème 1.11 (Charon, Honkala, Hudry, Lobstein [28])**.** *Pour tout $r\geq 2$, la borne inférieure $D_{r,1}$ des densités des codes $(r,\leq 1)$-identifiants dans la grille du roi est égale à*

$$D_{r,1}=\frac{1}{4r}.$$

Plus généralement, on définira $D_{r,\ell}$ comme la borne inférieure des densités des codes $(r,\leq\ell)$-identifiants dans la grille du roi. Notons que la grille du roi admet des $(r,\leq 3)$-jumeaux pour tout $r\geq 1$: on a toujours l'inclusion

$$B((0,0),r)\subset B((-1,0),r)\cup B((1,0),r)$$

et les ensembles $\{(-1,0),(1,0)\}$ et $\{(-1,0),(0,0),(1,0)\}$ ne peuvent donc pas être r-séparés.

Nous reproduisons la table dressée par M. Pelto dans [80], qui donne les meilleures bornes pour sur $D_{r,\ell}$ connues à ce jour (table 1.VI).

	$\ell = 1$	$\ell = 2$	$\ell \geq 3$
$r = 1$	$D_{1,1} = \frac{2}{9}$ [32, 40]	$\frac{5}{12} \leq D_{2,1} \leq \frac{3}{7}$ [80] [60]	n'existe pas [60]
$r = 2$	$D_{2,1} = \frac{1}{8}$ [28]	$\frac{31}{120} \leq D_{2,2} \leq \frac{2}{7}$ [60] [80]	n'existe pas [60]
$r \geq 3$	$D_{r,1} = \frac{1}{4r}$ [28]	$D_{r,2} = \frac{1}{4}$ [60]	n'existe pas [60]

Table 1.VI – *Valeurs connues à ce jour concernant les densités minimales $D_{r,\ell}$ des codes $(r, \leq \ell)$-identifiants dans la grille du roi*

Chapitre 2

Structure des graphes sans jumeaux

Sommaire

2.1 Réduction des graphes sans jumeaux 24
2.2 Paramètres extrémaux des graphes sans r-jumeaux 25
2.2.1 Ordre et taille 26
2.2.2 Degré minimal 26
2.3 Longues chaînes dans les graphes sans r-jumeaux 27
2.4 Structure des graphes sans $(1, \leq \ell)$-jumeaux 27
2.4.1 Graphes sans $(1, \leq 2)$-jumeaux 27
2.4.2 Cas général 29

Nous avons vu que la notion de codes $(r, \leq \ell)$-identifiants appelle immédiatement celle de *graphes sans* $(r, \leq \ell)$*-jumeaux*, qui sont précisément les graphes admettant de tels codes (voir 1.4). Rappelons que nous utilisons, comme dans toute cette partie, la convention 1.1 pour les notations.

2.1 Réduction des graphes sans jumeaux

Nous commençons ce chapitre par une propriété des graphes sans 1-jumeaux, qui étend un résultat de [36]. La preuve reprend l'idée utilisée par J. Moncel et S. Gravier dans [54].

Théorème 2.1. *Soit G un graphe connexe sans jumeaux ayant au moins 4 sommets. Il existe un sommet $v_0 \in V(G)$ tel que $G \setminus v_0$ soit aussi sans jumeaux et connexe.*

Preuve.
Soit $\mathscr{A}$ l'ensemble des sommets v de G tels que le graphe $G \setminus v$ soit connexe. L'ensemble $\mathscr{A}$ est non vide : par exemple, les extrémités d'un diamètre de G, ou d'une chaîne de G de longueur maximale, appartiennent à $\mathscr{A}$. Choisissons un sommet $a \in \mathscr{A}$ dont le degré soit le plus grand possible. Si $G \setminus a$ est sans jumeaux, on pose $v_0 = a$; sinon, c'est qu'il existe deux sommets distincts b_1 et b_2 de G tels que

$$N_G[b_1] = N_G[b_2] \cup \{a\}. \tag{2.1}$$

Dans ce cas, puisque tout sommet adjacent à b_2 l'est aussi à b_1, on a $b_2 \in \mathscr{A}$. Supposons donc à nouveau que $G \setminus b_2$ ait des jumeaux (sinon, nous pouvons poser $v_0 = b_2$) : il existe deux sommets distincts x et y dans G, tels que

$$N_G[x] = N_G[y] \cup \{b_2\}. \tag{2.2}$$

Puisque $x \in N_G[b_2]$, d'après (2.1) on a aussi $x \in N_G[b_1]$. De façon à avoir (2.2), on a nécessairement $y \in N_G[b_1] \setminus N_G[b_2]$ et donc $y = a$. Ainsi,

$$N_G[x] = N_G[a] \cup \{b_2\}. \tag{2.3}$$

Montrons que $x \in \mathscr{A}$, contredisant alors le choix de a par 2.3. Si $x = b_1$, par (2.1) et (2.3), les sommets n'appartenant pas à $\{b_1, b_2, a\}$ peuvent soit être adjacent, à ces trois sommets, soit à aucun des trois. Puisque $n \geq 4$, le premier cas se produit pour au moins un sommet de G : il est alors immédiat de vérifier que $G \setminus b_1$ est connexe, donc $b_1 \in \mathscr{A}$.

Si $x \neq b_1$, alors par (2.3) on a aussi $x \neq a$ et $x \neq b_2$. Ainsi, $x \in \mathscr{A}$ puisque les sommets adjacents à x sont tous adjacents à a, à part b_2 mais ce dernier est relié à a par la chaîne $b_2 b_1 a$. □

On peut aussi formuler ce résultat de la façon suivante, permettant de construire un graphe sans jumeaux en ajoutant les sommets un à un.

Corollaire 2.2. *Soit G un graphe connexe sans jumeaux ayant n sommets, $n \geq 4$. Alors il existe une suite de graphes connexes sans jumeaux*

$$G_3, G_4, \cdots, G_n$$

où G_3 est une chaîne à 3 sommets et $G_n = G$, telle que pour $3 \leq i \leq n-1$, le graphe G_i ait i sommets et soit un sous-graphe induit de G_{i+1}.

En combinant ce résultat avec le lemme suivant et le fait que $\gamma_1^{id}(\mathcal{P}_3) = 2$, on trouve une nouvelle preuve du fait qu'un code identifiant dans un graphe sans jumeaux d'ordre n nécessite au plus $n - 1$ sommets (théorème 1.8).

Lemme 2.3 (Charon, Hudry, Lobstein (communication personnelle))**.** *Soit G un graphe sans jumeaux connexe et $v \in V(G)$ tel que le graphe $G \setminus v$ soit aussi connexe et sans jumeaux. Alors*

$$\gamma^{id}(G) \leq \gamma^{id}(G \setminus v) + 1.$$

Preuve du lemme 2.3. Soit $\mathscr{C}$ un code identifiant de taille minimale de $G \setminus v$, et considérons-le comme un code de G. Etant donné que pour toute paire de sommets distincts x et y dans $V(G) \setminus \{v\}$, les sommets x et y sont couverts et séparés par $\mathscr{C}$, seul v peut, ou bien n'être pas couvert par $\mathscr{C}$, ou bien ne pas être séparé d'un unique sommet $v' \neq v$. Dans le premier cas, il suffit d'ajouter v à $\mathscr{C}$ pour faire de $\mathscr{C}$ un code identifiant de G, et dans le deuxième cas il suffit d'ajouter un sommet de $N_G[v] \Delta N_G[v']$, non vide par hypothèse. □

2.2 Paramètres extrémaux des graphes sans r-jumeaux

Ce travail avait été commencé par Charon, Honkala, Hudry et Lobstein dans [30] et [31] ; nous nous sommes joints à eux pour le poursuivre dans [6]. Nous ne présentons ici que les résultats concernant la taille et le degré minimal ; se reporter à l'article [6] (annexe A) pour les autres paramètres.

2.2.1 Ordre et taille

Rappelons que l'ordre d'un graphe G est son nombre de sommets (ici noté n) et sa taille est son nombre d'arêtes (ici noté ϵ). Le résultat qui suit peut être vu comme une conséquence des théorèmes 2.8 et 2.10 qui seront abordés dans la prochaine section.

Théorème 2.4. *Soit $r \geq 1$ et G un graphe connexe sans r-jumeaux d'ordre $n \geq 2$ et de taille ϵ. Alors $n \geq 2r+1$ et $\epsilon \geq 2r$, et le seul de ces graphes admettant $2r+1$ sommets ou $2r$ arêtes est la chaîne P_{2r+1}.*

Notons qu'aucun autre paramètre n'étant contraint, il existe des graphes sans r-jumeaux connexes ayant un ordre ou une taille arbitrairement grands ; en revanche, trouver la valeur maximale de la taille ϵ devient plus difficile quand l'ordre n du graphe est fixé. Commençons par le cas $r = 1$:

Théorème 2.5 (Auger, Charon, Honkala, Hudry, Lobstein [6] & annexe A)**.** *Soit G un graphe connexe sans 1-jumeaux d'ordre $n \geq 3$ et de taille ϵ. Alors*

$$\epsilon \leq \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor,$$

et le seul de ces graphes satisfaisant l'égalité est le graphe hyperoctaédral $\mathcal{H}_n$.

Pour $r > 1$, le problème est plus délicat et les résultats précis complexes à énoncer ; nous renvoyons à l'article [6] (annexe A) pour les détails. Néanmoins, nous pouvons donner l'ordre de grandeur asymptotique suivant :

Théorème 2.6 (Auger, Charon, Honkala, Hudry, Lobstein [6] & annexe A)**.** *Soit $r \geq 3$ fixé et soit $F_{r,n}(\epsilon)$ la taille maximum d'un graphe connexe sans r-jumeaux d'ordre n. Alors*

$$.63(r - .915) n \log_2 n \precsim \binom{n}{2} - F_{r,n}(\epsilon) \precsim rn \log_2 n,$$

où la notation $f(n) \precsim g(n)$ signifie

$$\limsup_{n\to\infty} \frac{g(n)}{f(n)} \geq 1.$$

2.2.2 Degré minimal

Il est clair qu'il existe pour toutes valeurs de n et r (avec $n \geq 2r+1$) des graphes sans r-jumeaux connexes de degré minimal 1 : une simple chaîne convient. Une question plus intéressante est de déterminer la valeur *maximale* du degré minimal $\Delta_{\min}$ d'un tel graphe. Voici un des résultats :

Théorème 2.7 (Auger, Charon, Honkala, Hudry, Lobstein [6] & annexe A)**.** *Soit G un graphe connexe sans r-jumeaux d'ordre $n \geq 2$ et de degré minimal $\Delta_{\min} \geq 2$. Pour tout $r \geq 6$ on a*

$$\Delta_{\min} \leq \min\left\{ \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1, \frac{3n - r + 2}{2r - 5} \right\}.$$

2.3 Longues chaînes dans les graphes sans r-jumeaux

Deux sommets distincts x et y d'un graphe sans r-jumeaux doivent être r-séparés, d'où l'existence d'un sommet dont la distance à x, ou à y, est au moins $r+1$. Ceci laisse présager l'existence de longues chaînes dans un tel graphe ; et en effet la propriété suivante a été établie :

Théorème 2.8 (Charon, Hudry, Lobstein [29])**.** *Pour tout $r \geq 1$, un graphe sans r-jumeaux d'ordre au moins 3 contient une chaîne à $2r+1$ sommets en tant que sous-graphe partiel.*

Les auteurs ont conjecturé dans [29] qu'il existerait en fait une telle chaîne en tant que sous-graphe *induit* dans un graphe sans r-jumeaux, c'est-à-dire un chaîne de la forme

$$x_1x_2\cdots x_{2r+1}$$

où deux sommets x_i et x_j ne peuvent être adjacents dès que $|i-j|>1$.

Nous avons prouvé cette conjecture dans [5], en reprenant l'idée d'une preuve due à F. Chung dans l'article [46].

Théorème 2.9 (Auger [5])**.** *Un graphe de rayon $\rho \geq 1$ et contenant un centre c dont aucun voisin n'est un centre admet une chaîne à $2\rho+1$ sommets en tant que sous-graphe induit.*

De ce résultat, nous déduisons facilement la conjecture :

Théorème 2.10. *Pour tout $r \geq 1$, un graphe sans r-jumeaux d'ordre au moins 3 contient une chaîne à $2r+1$ sommets en tant que sous-graphe induit.*

Un corollaire intéressant du théorème 2.10 est le suivant :

Corollaire 2.11. *Pour tout $r \geq 1$, la chaîne P_{2r+1} à $2r+1$ sommets est le plus petit graphe sans r-jumeaux d'ordre au moins 3.*

Nous avons volontairement laissé une ambiguïté ci-dessus dans le superlatif «le plus petit» : il peut s'entendre au sens du nombre de sommets, ou des relations d'ordre «être un sous-graphe induit de» ou «être un sous-graphe partiel de» ; pour ces différentes relations la chaîne P_{2r+1} est un élément minimal de l'ensemble des graphes sans r-jumeaux d'ordre au moins 3.

Une autre preuve du théorème [5], n'utilisant pas [46] mais plus complexe, peut aussi être consultée dans [4].

2.4 Structure des graphes sans $(1, \leq \ell)$-jumeaux

2.4.1 Graphes sans $(1, \leq 2)$-jumeaux

Il est immédiat de constater qu'un graphe G sans $(1, \leq 2)$-jumeaux ne peut posséder de sommet de degré 1 : en effet, si a est un sommet de degré 1 ayant pour unique voisin un sommet b, on a toujours

$$N_G[b] = N_G[\{a,b\}].$$

Il s'ensuit qu'un graphe sans $(1, \leq 2)$-jumeaux d'ordre au moins 3 possède nécessairement un cycle. Nous avons prouvé le résultat suivant :

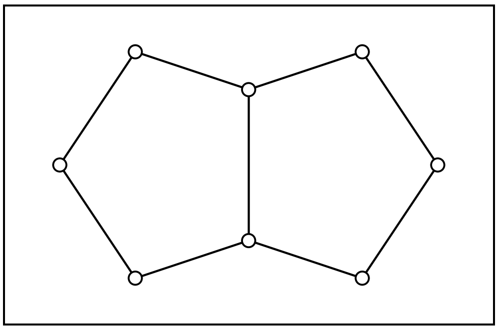


Figure 2.I – *Un graphe sans* $(1, \leq 2)$*-jumeaux et dont le plus grand cycle sans corde est de longueur 5*

Théorème 2.12 (Auger, Charon, Hudry, Lobstein [14], [9] & annexe C)**.** *Un graphe sans* $(1, \leq 2)$*-jumeaux d'ordre au moins* 3 *contient un cycle de longueur au moins* 7 *en tant que sous-graphe partiel.*

Puisque le plus petit cycle sans $(1, \leq 2)$-jumeaux compte 7 sommets, ce résultat est l'analogue des théorèmes 2.8 et 2.10 dans le cas $(r, \ell) = (1, 2)$. Le plus petit cycle sans $(r, \leq 2)$-jumeaux étant le cycle de longueur $4r + 3$, il n'y a de là qu'un pas pour conjecturer ce qui suit :

Conjecture 2.13. *Pour tout* $r \geq 1$*, un graphe sans* $(r, \leq 2)$*-jumeaux d'ordre au moins* 3 *contient un cycle de longueur au moins* $4r + 3$ *en tant que sous-graphe partiel.*

Cependant, contrairement au cas du théorème 2.10, un tel cycle ne saurait toujours exister en tant que sous-graphe induit, en effet déjà pour $r = 1$ on peut trouver un graphe sans $(1, \leq 2)$-jumeaux, ayant plus d'une arête, dont le plus grand cycle sans corde est de longueur 5 (voir sur la figure 2.I).

Quant à la structure des graphes sans $(1, \leq 3)$-jumeaux, peu de choses sont connues à l'heure actuelle. Le plus petit de ces graphes actuellement connu a été donné par T. Laihonen dans [68] et est représenté sur la figure 2.II. Ce graphe possède 16 sommets et tous ses sommets sont de degré 3. Il paraît cependant prématuré de penser que tous les graphes sans $(1, \leq 3)$-jumeaux pourraient contenir un sous-graphe apparenté à celui-ci. La généralisation des théorèmes 2.10 et 2.12 au cas $\ell \geq 3$ est donc encore à imaginer.

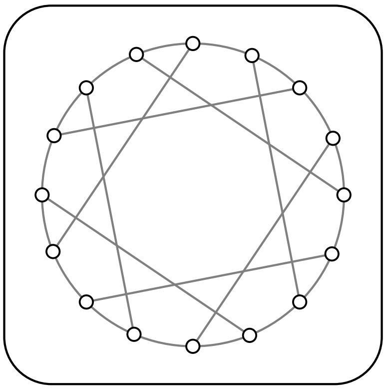


Figure 2.II – *Un graphe sans* $(1, \leq 3)$*-jumeaux.*

2.4.2 Cas général

On en sait encore peu au sujet des graphes sans $(r, \leq \ell)$-jumeaux, et ce même quand $r = 1$. La propriété suivante est due à J. Moncel et T. Laihonen.

Théorème 2.14 (Moncel, Laihonen [69])**.** *Soient $\ell \geq 1$ et G un graphe sans $(1, \leq \ell)$-jumeaux, connexe, ayant au moins une arête. Alors pour tout sommet $v \in V(G)$, le voisinage ouvert $N(v)$ contient un stable de cardinal ℓ.*

En particulier, le degré minimal d'un graphe sans $(1, \leq \ell)$-jumeaux est supérieur ou égal à ℓ, ce qui a déjà été observé dans [75]. Un autre résultat permet à l'inverse d'établir qu'un graphe est sans $(1, \leq \ell)$-jumeaux :

Théorème 2.15 (Moncel, Laihonen [69])**.** *Si $\ell \geq 1$, les graphes ℓ-réguliers de maille supérieure ou égale à 7, ainsi que les graphes $(\ell + 1)$-réguliers de maille supérieure ou égale à 5, sont sans $(1, \leq \ell)$-jumeaux.*

Prouvons un résultat recouvrant partiellement ceux qui précèdent :

Théorème 2.16. *Soit $\ell \geq 1$.*

- *Si G est un graphe connexe sans $(1, \leq \ell)$-jumeaux connexe d'ordre au moins 3 et si $x \in V(G)$, alors toute partie $A \subset V(G) \setminus \{x\}$ couvrant $N_G[x]$ est de cardinal supérieur ou égal à ℓ ;*
- *réciproquement, si G est un graphe tel que pour tout sommet $x \in V(G)$, toute partie $A \subset V(G) \setminus \{x\}$ couvrant $N_G[x]$ soit de cardinal supérieur ou égal à $\ell + 1$, alors G est sans $(1, \leq \ell)$-jumeaux.*

Preuve. Soit G sans $(1, \leq \ell)$-jumeaux, connexe, comportant au moins deux sommets et soit $A \subset V(G) \setminus \{x\}$ couvrant $N_G[x]$ où $x \in V(G)$. Alors $N_G[A] = N_G[A \cup \{x\}]$ donc $|A \cup \{x\}| \geq \ell + 1$. Réciproquement, soit G un graphe vérifiant l'hypothèse de la seconde partie du théorème, et soient X et Y deux ensembles d'au plus ℓ sommets de G, tels que

$$N_G[X] = N_G[Y].$$

Si X et Y sont distincts, sans perte de généralité on peut supposer qu'il existe $x \in X \setminus Y$. Alors $N_G[x]$ est couvert par la partie Y, contredisant notre hypothèse. □

On retrouve aisément la seconde partie du théorème 2.15. D'autre part, un stable maximum de $N_G(x)$ couvre nécessairement $N_G[x]$ et on retrouve donc le theorème 2.14.

Chapitre 3

Complexité algorithmique et codes identifiants

Sommaire

3.1 Introduction et état des lieux 30
3.2 Des résultats de $\mathcal{NP}$-complétude 31
3.3 Un algorithme linéaire pour les arbres 32
3.3.1 Principe de l'algorithme 32
3.3.2 Codes presque identifiants 34
3.3.3 Généralisation aux graphes de largeur d'arbre bornée 35

3.1 Introduction et état des lieux

Dans ce chapitre, nous tentons d'évaluer la complexité algorithmique des problèmes de calcul de codes identifiants dans les graphes. Est-il envisageable de programmer un ordinateur pour calculer un code identifiant optimal dans un graphe à 1000 sommets ? Serait-ce plus simple si le graphe avait une structure particulière, par exemple s'il était un arbre ? Les différents problèmes que nous allons envisager peuvent être décrits par le problème de décision générique suivant, qui dépend de deux entiers $r \geq 1$, $\ell \geq 1$ et d'une classe $\mathscr{H}$ de graphes :

Min $(r, \leq \ell)$-IdCode dans $\mathscr{H}$
Instance : Un graphe $G \in \mathscr{H}$ et un entier k ;
Question : G admet-il un code $(r, \leq \ell)$-identifiant de taille au plus k ?

Le terme IdCode est bien entendu une abréviation pour *code identifiant*, et comme d'habitude nous omettrons les références superfétatoires à r ou ℓ dans les cas où ils sont égaux à 1 (convention 1.1), ainsi que celle à $\mathscr{H}$ si la classe en question est celle de tous les graphes. Après un rappel des résultats déjà connus quant aux problèmes relevant de ce type, nous présenterons de nouveaux résultats de $\mathcal{NP}$-complétude pour différentes valeurs de r, ℓ et différentes classes $\mathscr{H}$ de graphes. Ceci fait, nous exhiberons un algorithme linéaire permettant de calculer efficacement un code identifiant de taille minimale dans un arbre donné. Nous renvoyons à [50] pour les définitions des notions relatives à la complexité algorithmique utilisées dans ce chapitre.

Les résultats précédemment établis sur la complexité algorithmique des problèmes de minimisation des codes identifiants concernent tous le cas $\ell = 1$; à notre connaissance un seul [34] aborde le cas $r > 1$.

Théorème 3.1 (Cohen, Honkala, Lobstein, Zémor [39], Charon, Hudry, Lobstein [34]). *Pour tout $r \geq 1$, le problème* Min r-IdCode *est $\mathcal{NP}$-complet.*

La recherche d'un code r-identifiant de taille minimale dans un graphe donné est donc un problème algorithmiquement difficile. Sous réserve de validité de l'hypothèse $\mathcal{P} \neq \mathcal{NP}$, qui est au moins satisfaite dans la pratique jusqu'à preuve du contraire, un algorithme qui calculerait un code r-identifiant de taille minimale dans tout graphe donné ne pourrait fonctionner qu'en un temps *exponentiel* (dans le pire des cas) par rapport à la taille du graphe.

Quand un problème est $\mathcal{NP}$-complet, il est d'usage d'étudier les *algorithmes polynomiaux d'approximation* pour ce problème. Un tel algorithme doit fonctionner en temps polynomial par rapport à la taille du graphe et retourner un code identifiant ; on mesure la «marge d'erreur» de cet algorithme par son *rapport d'approximation* qui est le quotient, dans le pire des cas, entre la taille du code identifiant retourné par l'algorithme et la taille optimale d'un code identifiant dans le graphe. Ainsi, ce rapport est d'autant plus proche de 1 que l'algorithme est efficace.

Dans le cas $r = 1$, Gravier, Moncel et Klasing sont allés plus loin que le théorème 3.1 en prouvant que Min IdCode est APX-difficile ([52]). Rappelons que ceci implique, sous réserve de l'hypothèse $\mathcal{P} \neq \mathcal{NP}$, que Min IdCode n'admet pas d'algorithmes d'approximation de rapport constant en dessous d'une certaine valeur $c > 1$. Ainsi, on ne saurait concevoir des algorithmes efficaces avec une marge d'erreur arbitrairement petite.

Avec une hypothèse plus forte, Laifenfeld et Trachtenberg ([67]) ont prouvé que Min IdCode n'admettait pas d'algorithme polynomial d'approximation dont le rapport d'approximation serait en deçà de $(1 - \epsilon) \log n$, où n est l'ordre du graphe, et ce pour tout $\epsilon > 0$. L'hypothèse en question est la non-inclusion suivante :

$$\mathcal{NP} \nsubseteq DTIME(n^{\log \log n}).$$

Si cette hypothèse était fausse, elle signifierait que tous les problèmes de la classe $\mathcal{NP}$ admettent un algorithme déterministe les résolvant en temps $O(n^{\log \log n})$. On voit donc que cette hypothèse est plus forte que l'hypothèse $\mathcal{P} \neq \mathcal{NP}$.

Finalement, différents auteurs ont remarqué qu'un algorithme glouton appliqué au problème Min r-IdCode fournissait un algorithme polynomial d'approximation de rapport d'approximation $O(\log n)$ ([67], [88]).

3.2 Des résultats de $\mathcal{NP}$-complétude

Pour pousser l'analyse de la complexité au-delà de la constatation de $\mathcal{NP}$-complétude, nous avons choisi une autre direction que celle qui a été empruntée dans [52] et [67] : nous étudions la complexité de Min $(r, \leq \ell)$-IdCode dans $\mathscr{H}$ pour différentes classes $\mathscr{H}$ de graphes. Les classes qui nous concernent sont les suivantes :

- la classe $\mathscr{P}_{\Delta \leq 3}$ des graphes planaires de degré maximal inférieur ou égal à 3 ;
- pour tout $k \geq 3$, la classe $\mathscr{P}^{m \geq k}_{\Delta \leq 4}$ des graphes planaires de degré maximal au plus 4 et dont la maille est au moins k ;
- la classe $\mathscr{F}$ des forêts.

Commençons par un résultat valable pour $\ell = 1$ et $\ell = 2$:

Théorème 3.2 (Auger, Charon, Hudry, Lobstein [7] & annexe D)**.** *Pour tout* $r \geq 1$ *et* $\ell \in \{1, 2\}$*, le problème* Min $(r, \leq \ell)$-IdCode dans $\mathscr{P}_{\Delta \leq 3}$ *est* $\mathcal{NP}$*-complet.*

Ce résultat étend ainsi le théorème 3.1 en prouvant que même le fait de limiter les instances à la classe $\mathscr{P}_{\Delta \leq 3}$, pourtant beaucoup plus restreinte, fait toujours de la recherche d'un code r-identifiant de taille minimale un problème $\mathcal{NP}$-difficile. Il constitue aussi à notre connaissance, dans le cas $\ell = 2$, le premier résultat concernant la complexité algorithmique de la recherche d'un code $(r, \leq \ell)$-identifiant optimal quand $\ell > 1$ (sujet ayant par ailleurs fait l'objet de nombreuses recherches d'un point de vue combinatoire, voir le chapitre 1). On pourrait d'ailleurs se contenter d'énoncer le corollaire suivant :

Corollaire 3.3. *Pour tout* $r \geq 1$*, le problème* Min $(r, \leq 2)$-IdCode *est* $\mathcal{NP}$*-complet.*

Dans le cadre de notre recherche d'un algorithme efficace dans la classe des arbres, nous avons établi une variante de ce résultat dans le cas $r = 1$ et $\ell = 1$, en nous intéressant à la classe $\mathscr{P}^{m \geq k}_{\Delta \leq 4}$ des graphes planaires de maille au moins k et de degré maximal au plus 4 :

Théorème 3.4 (Auger [3] & annexe E)**.** *Pour tout* $k \geq 3$*, le problème* Min IdCode dans $\mathscr{P}^{m \geq k}_{\Delta \leq 4}$ *est* $\mathcal{NP}$*-complet.*

La raison de l'étude de ces classes de graphes est qu'elles constituent des approximations, en un certain sens, de la classe des arbres, ces derniers pouvant être considérés comme ayant une maille infinie. Plus précisément, ces classes forment une suite décroissante au sens de l'inclusion, et leur intersection pour $k \geq 3$ est la classe des graphes planaires de degré au plus 4 ne contenant aucun cycle : une sous-classe de celle des *forêts*. Or nous verrons dans la prochaine section un algorithme qui résout le problème Min IdCode dans $\mathscr{F}$ en un temps linéaire, donc polynomial, en l'ordre du graphe. Autour de cette classe, nous disposons de l'infinité des classes $\mathscr{P}^{m \geq k}_{\Delta \leq 4}$, en un certain sens aussi proches que l'on veut de $\mathscr{F}$, où le problème de minimisation reste $\mathcal{NP}$-complet.

Avant de terminer cette partie, notons qu'il n'existe pour l'instant aucun résultat sur la complexité algorithmique des problèmes Min $(r, \leq \ell)$-IdCode lorsque $\ell \geq 3$, et ce même pour $r = 1$. Toutefois, il serait très surprenant que ces problèmes soient faciles et ce n'est pas faire preuve d'une bien grande témérité que de risquer ce qui suit :

Conjecture 3.5. *Pour tout* $r \geq 1$ *et tout* $\ell \geq 3$*, le problème* Min $(r, \leq \ell)$-IdCode *est* $\mathcal{NP}$*-complet.*

3.3 Un algorithme linéaire pour les arbres

3.3.1 Principe de l'algorithme

Nous présentons ici un algorithme permettant de calculer un code identifiant de taille minimale dans un arbre ou une forêt donnés, en un temps linéaire par rapport à l'ordre du graphe. On se limitera sans perte de généralité à la description de l'algorithme lorsque le graphe donné en instance est un arbre. Le détail de l'algorithme peut être trouvé dans [3] (annexe E), nous allons juste ici en expliquer le principe. Considérons un arbre $\mathcal{A}$ et une arête v_1v_2 de $\mathcal{A}$, nous voulons procéder de manière récursive comme illustré sur la figure 3.I :

- on supprime l'arête v_1v_2 de $\mathcal{A}$, obtenant ainsi deux arbres $\mathcal{A}_1$ et $\mathcal{A}_2$ contenant respectivement les sommets v_1 et v_2 ;
- on applique récursivement l'algorithme dans les sous-arbres $\mathcal{A}_1$ et $\mathcal{A}_2$, obtenant pour chacun d'eux un code identifiant ;

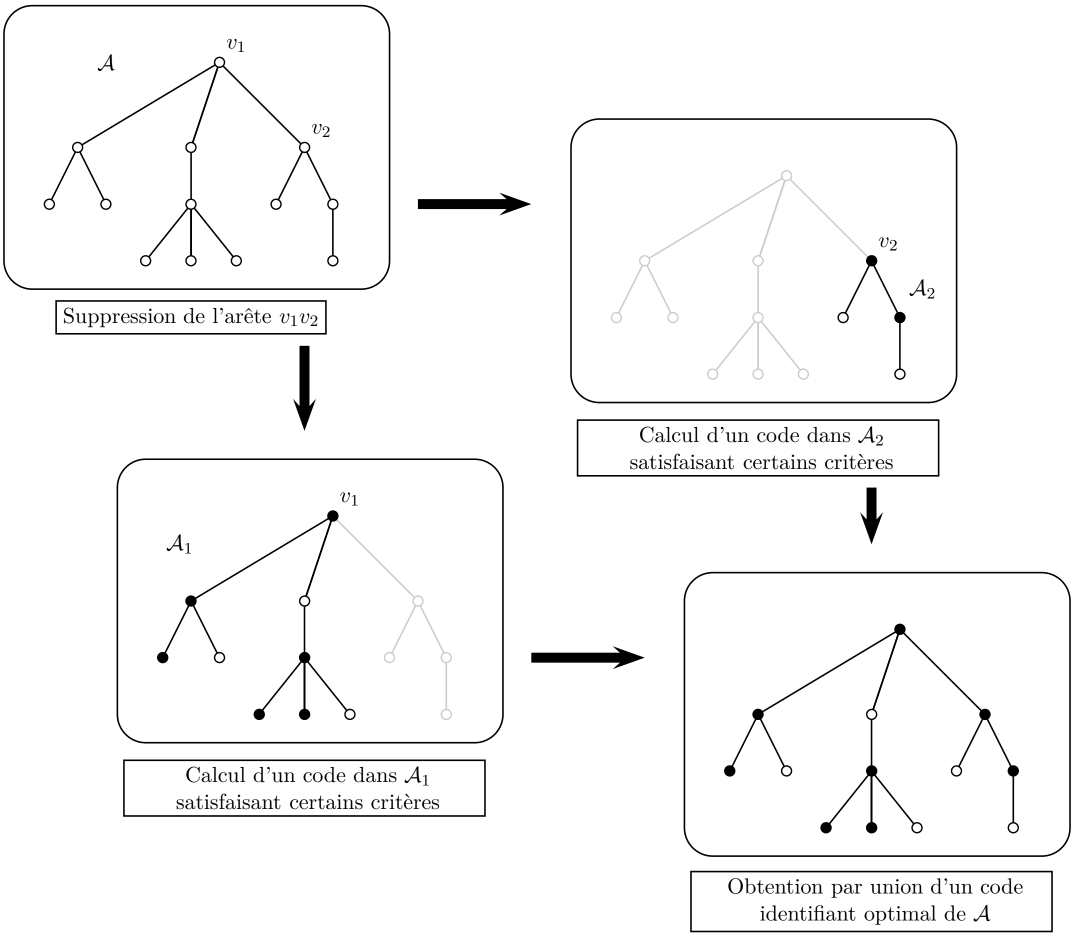


Figure 3.I – *Principe de l'algorithme de calcul d'un code identifiant optimal dans un arbre. Les sommets noirs sont les mots de code.*

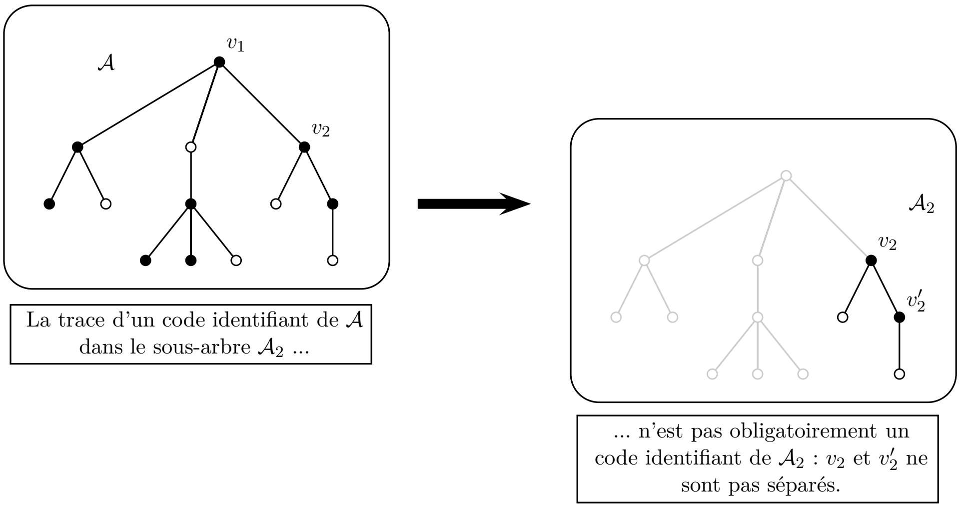


Figure 3.II – *Un problème pour la récursivité*

– on calcule l'union des deux codes obtenus, espérant ainsi obtenir un code identifiant pour $\mathcal{A}$.

Bien entendu, appliqué tel quel, l'algorithme ne peut fonctionner et comme indiqué sur la figure 3.I ce ne sont pas des codes identifiants que l'on va chercher dans $\mathcal{A}_1$ et $\mathcal{A}_2$ mais des codes satisfaisant *certains critères*, que nous allons maintenant détailler. Le principal problème que nous rencontrons, outre le souci de minimalité, est qu'en considérant au départ un code identifiant de $\mathcal{A}$, les codes obtenus dans les sous-arbres $\mathcal{A}_1$ et $\mathcal{A}_2$ après suppression de v_1v_2 ne sont pas nécessairement des codes identifiants, comme illustré sur la figure 3.II.

3.3.2 Codes presque identifiants

Ainsi, en gardant les notations du paragraphe précédent, un code identifiant pour $\mathcal{A}$, s'il était l'unique code identifiant minimal, ne saurait être obtenu directement comme union de codes identifiants de $\mathcal{A}_1$ et $\mathcal{A}_2$ et il nous faut introduire une notion plus souple que celle de code identifiant :

Définition 3.6. *Si G est un graphe et v un sommet de G, un code v-presque identifiant est un sous-ensemble $\mathscr{C}$ de $V(G)$ tel que les sous-ensembles*

$$N_G[x] \cap \mathscr{C}$$

pour tout $x \in V(G) \setminus \{v\}$ soient non vides et distincts.

Ainsi dans un code v-presque identifiant, le sommet v peut être choisi comme mot de code et donc utilisé pour couvrir et séparer, mais nous n'avons pas à identifier v. Cette fois-ci, on peut vérifier facilement que si $\mathscr{C}$ est un code v_1-presque identifiant de $\mathcal{A}$, alors les codes $\mathscr{C} \cap V(\mathcal{A}_1)$ et $\mathscr{C} \cap V(\mathcal{A}_2)$ sont respectivement des codes v_1-presque identifiant de $\mathcal{A}_1$ et v_2-presque identifiant de$\mathcal{A}_2$.

Pour assurer la réciproque et calculer un code minimal, il va nous falloir en fait calculer plusieurs types de codes presque identifiants à chaque étape, 10 pour être précis, satisfaisant différentes combinaisons des propriétés suivantes, par exemple pour un code v_1-presque identifiant :

- on peut exiger que le code soit en fait identifiant ;
- on peut exiger que v_1 lui-même soit un mot de code ;
- on peut exiger que v_1 soit adjacent à au moins un mot de code ;
- on peut enfin exiger qu'il existe un *voisin préféré* de v_1, c'est-à-dire un sommet adjacent à v_1 qui ne soit couvert que par le mot de code v_1.

Ces propriétés sont suffisantes pour établir des formules de récurrence entre les tailles minimales de ces différents codes dans $\mathcal{A}$, $\mathcal{A}_1$ et $\mathcal{A}_2$, nous fournissant un algorithme qui fonctionne en un temps linéaire par rapport à la taille de l'arbre passé en instance. Nous renvoyons à [3] (annexe E) pour les détails de l'algorithme.

3.3.3 Généralisation aux graphes de largeur d'arbre bornée

Dans sa thèse [74], Moncel a remarqué qu'un théorème assez général dû à Arnborg, Lagergren et Seese et concernant l'algorithmique dans la classe des graphes de largeur d'arbre bornée ([2]), impliquait plus généralement :

Théorème 3.7 (Moncel [74])**.** *Pour tout $k \geq 1$, il existe un algorithme qui calcule en temps linéaire un code identifiant de taille minimale pour tout graphe de largeur d'arbre au plus k.*

Pour la notion de largeur d'arbre, nous renvoyons (par exemple) à [21]. Contentons-nous de dire que les forêts ont précisément une largeur d'arbre 1, et que ce théorème implique donc l'existence d'un algorithme comme le nôtre. Ceci dit, le théorème 3.7 donne l'existence théorique d'un algorithme, et ne permet pas de calculer effectivement un code identifiant.

Chapitre 4

Systèmes de contrôle

Sommaire

4.1 Introduction **36**
4.2 Définitions **37**
4.3 Premières propriétés **39**
4.3.1 Lien avec les codes identifiants 39
4.3.2 Existence de systèmes de contrôles 39
4.3.3 Lien avec les ensembles dominants 39
4.3.4 Lien avec les codes superposés 40
4.3.5 Propriétés de monotonie 40
4.3.6 Compression 40
4.3.7 Complexité algorithmique 41
4.4 Bornes en fonction du nombre de sommets **41**
4.4.1 Bornes inférieures 41
4.4.2 Bornes supérieures 41
4.5 Systèmes de contrôle dans les chaînes et les cycles **43**
4.6 Systèmes de contrôle dans la grille du roi **44**
4.6.1 Identification des sommets 45
4.6.2 Identification des ensembles de sommets 45

4.1 Introduction

L'idée à la base des systèmes de contrôle dans les graphes est d'étendre la notion de code identifiant en apportant un peu de souplesse au rôle des mots de code, en leur offrant en particulier la possiblité de ne pas couvrir un sommet pourtant à portée de l'être. Nous allons voir que, paradoxalement, donner moins de sommets à couvrir à chacun des mots de code peut amener à des codes bien plus efficaces. Expliquons-nous grâce à un nouvel exemple de musée, que l'on pourra trouver sur la figure 4.I. À ce musée est associé un graphe G, construit en utilisant le même principe que dans le chapitre 1. Si l'on voulait surveiller ce musée (pour détecter du feu, du mouvement, ou autre) avec un système modélisé par un code identifiant, nous aurions besoin de cinq détecteurs, car G n'est autre que le graphe hyperoctaédral d'ordre 6 (voir 0.8), qui nécessite de considérer tous les sommets comme des mots de code, sauf un seul, afin de former un code identifiant.

Faisons maintenant l'hypothèse suivante : *on peut décider d'exclure certaines pièces de la zone de couverture des détecteurs.* En orientant les détecteurs d'une certaine manière, en

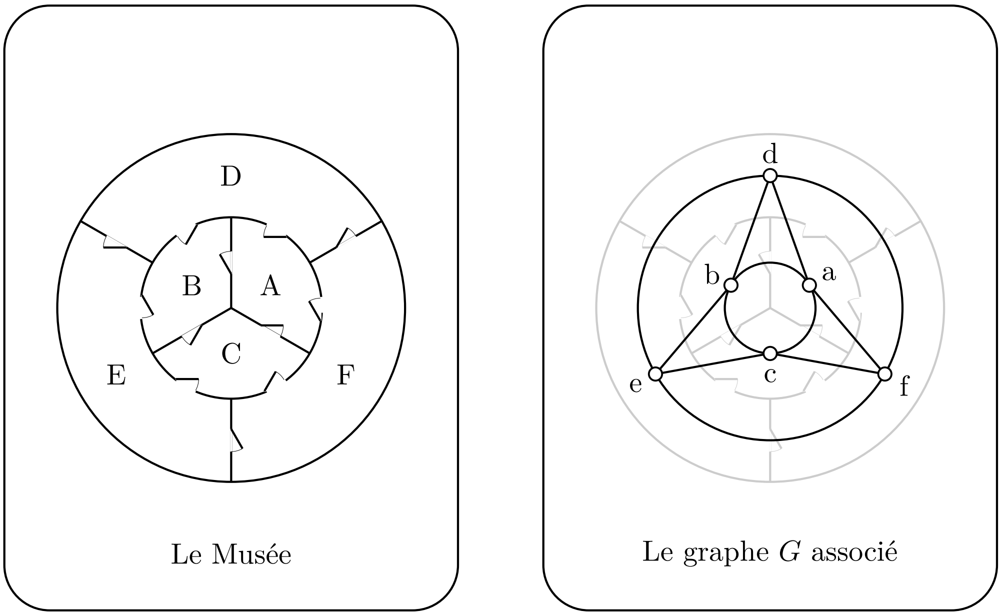


Figure 4.I – *Un nouveau musée à surveiller, et le graphe G associé.*

coupant certains circuits ou autre manipulation, on suppose donc que l'on peut assigner au détecteur une zone de contrôle plus petite que l'ensemble des pièces qui seraient à portée. Dans notre modèle mathématique, ceci signifie que l'on peut assigner à chaque mot de code c un ensemble de sommets $Z(c) \subseteq N_G[c]$ qui seront couverts par c. Les sommets de $N_G[c] \setminus Z(c)$ sont considérés comme non couverts par c, et pour le reste tout fonctionne comme pour un code identifiant. Nous parlerons alors de *contrôleurs* au lieu de mots de codes, et de *contrôle* au lieu de couverture. L'ensemble des contrôleurs qui ont en charge un sommet x s'appelle l'*étiquette* de x. La figure 4.II montre une manière optimale de procéder dans notre musée, utilisant trois contrôleurs, au lieu des cinq nécessaires pour un code identifiant. Dans cet exemple, tout se passe en fait comme si l'on était en présence d'un code identifiant sur un cycle de longueur 6 : on n'a pas tenu compte des arêtes qui ne nous intéressaient pas.

Nous nous autorisons, dans le cas des systèmes de contrôle, à placer plusieurs contrôleurs au même endroit, avec des zones de contrôle différentes. On pourrait considérer qu'il s'agit d'un seul détecteur avec plusieurs zones différentes à observer ; mais un tel équipement aurait besoin de plus de bits pour transmettre ses observations, et c'est ce nombre de bits (un par zone) que nous prenons en compte. Considérons un deuxième exemple, là encore un graphe extrémal où tous les sommets sauf un seraient nécessaires pour un code identifiant : il s'agit de l'étoile à 15 sommets $K_{1,14}$. La figure 4.III présente pour ce graphe un code identifiant optimal de taille 14, et un système de contrôle optimal de taille 4.

Comme pour les codes identifiants, nous étendrons les systèmes de contrôle à l'identification à distance d'ensembles de sommets et définirons ainsi les systèmes de $(r, \leq \ell)$-contrôle.

4.2 Définitions

Dans tout le chapitre, r et ℓ sont des entiers strictement positifs et G un graphe non orienté connexe. Un r-contrôleur de G est un couple

$$w = (c, Z)$$

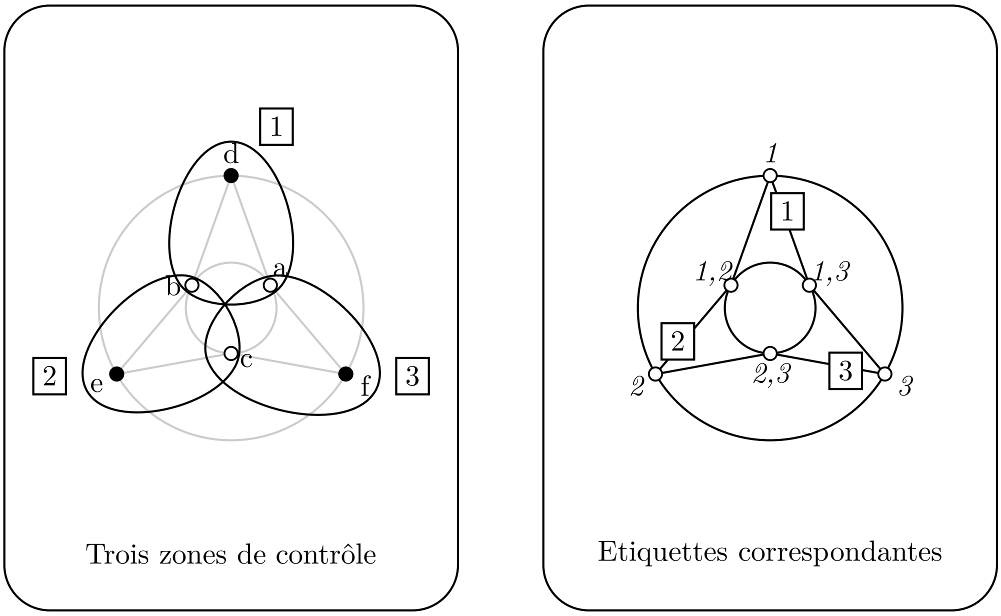


Figure 4.II – *Un système de contrôle optimal dans* G*. À gauche, les trois zones de contrôle assignées aux contrôleurs* 1, 2 *et* 3 *respectivement placés en* d*,* e *et* f*. À droite, le même système de contrôle représenté avec les notations que nous utiliserons dans cette partie : les localisations des contrôleurs* 1, 2 *et* 3 *sont indiquées par des boîtes, et on indique près de chaque sommet son étiquette, c'est-à-dire la liste des contrôleurs qui le contrôlent.*

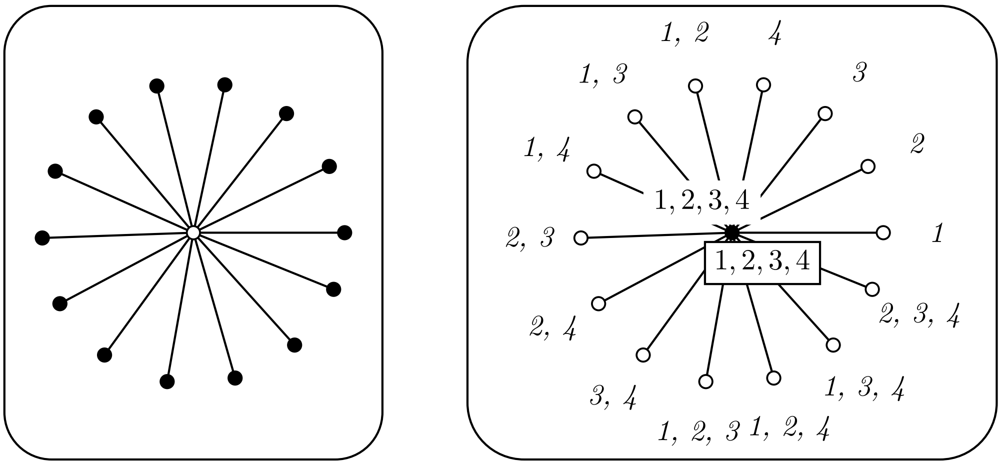


Figure 4.III – *À gauche, un code identifiant optimal (sommets noirs) sur l'étoile à 15 sommets. À droite, un système de contrôle optimal sur ce même graphe, avec le même système de notation que sur la figure 4.II.*

où c est un sommet de G et Z un sous-ensemble de $B_G(c, r)$. Nous disons que le r-contrôleur w est *localisé* en c, ou est *en* c, et que c est le *centre de contrôle* de w. L'ensemble Z est la *zone de contrôle* de w. Etant donné un r-contrôleur w, on notera respectivement $c(w)$ et $Z(w)$ le centre de contrôle et la zone de contrôle de w. Si x est un sommet de G et $x \in Z(w)$, nous disons que w *contrôle* x.

Si $\mathcal{W}$ est un ensemble (fini) de r-contrôleurs de G, nous définissons l'*étiquette* d'un sommet x de G dans $\mathcal{W}$ comme l'ensemble des contrôleurs de $\mathcal{W}$ qui contrôlent x. Nous notons

$$L_{\mathcal{W}}(v) = \{w \in \mathcal{W} : x \in Z(w)\}$$

cet ensemble. Si A est un ensemble de sommets de G, on pose alors

$$L_{\mathcal{W}}(A) = \bigcup_{x \in A} L_{\mathcal{W}}(x);$$

ainsi $L_{\mathcal{W}}(\emptyset) = \emptyset$. Un ensemble de r-contrôleurs $\mathcal{W}$ est alors dit *système de* $(r, \leq \ell)$-*contrôle* de G si les étiquettes $L_{\mathcal{W}}(A)$, pour toutes les parties A de $V(G)$ de taille au plus ℓ, sont deux à deux distinctes. Comme dans le cas des codes identifiants, si $\mathcal{W}$ est un ensemble de contrôleurs dans G, nous dirons qu'un sommet $x \in V(G)$ est *identifié* par $\mathcal{W}$ si $L_{\mathcal{W}}(x) \neq L_{\mathcal{W}}(y)$ pour tous les $y \in V(G)$ distincts de x.

Nous notons $w_{r,\ell}(G)$ la taille minimale d'un système de $(r, \leq \ell)$-contrôle dans un graphe G ; on abréviera $w_{r,1}(G)$ en $w_r(G)$.

4.3 Premières propriétés

4.3.1 Lien avec les codes identifiants

Un code $(r, \leq \ell)$-identifiant $\mathscr{C}$ dans un graphe sans $(r, \leq \ell)$-jumeaux G définit de manière naturelle un système de $(r, \leq \ell)$-contrôle de G, en associant à chaque mot de code $c \in \mathscr{C}$ le contrôleur $(c, B_G(c, r))$. Ainsi, quel que soit le graphe sans $(r, \leq \ell)$-jumeaux G considéré, on a l'inégalité

$$w_{r,\ell}(G) \leq \gamma^{id}_{r,\ell}(G).$$

En ceci, les systèmes de contrôle constituent bien une extension des codes identifiants.

4.3.2 Existence de systèmes de contrôles

Contrairement aux cas des codes identifiants, tout graphe admet des systèmes de $(r, \leq \ell)$-contrôle pour toutes les valeurs de r et ℓ. En particulier, dans tout graphe G l'ensemble des contrôleurs de la forme $(x, \{x\})$ pour tous les sommets x de G est un système de $(r, \leq \ell)$-contrôle pour tous $r \geq 1$ et $\ell \geq 1$; nous appelons *trivial* ce système de contrôle, et par extension tout système de contrôle contenant autant ou plus de contrôleurs que de sommets à contrôler.

4.3.3 Lien avec les ensembles dominants

Si $\mathcal{D}$ est un ensemble dominant d'un graphe G, on peut aisément construire à partir de $\mathcal{D}$ un système de $(1, \leq 1)$-contrôle pour G : il suffit de placer suffisamment de contrôleurs sur chaque élément x de $\mathcal{D}$ pour qu'il puisse identifier les sommets de $N_G[x]$, et tous les sommets de G seront alors identifiés. Il suffit clairement de $\lceil \log(k+1) \rceil$ pour identifier par des 1-contrôleurs une étoile de taille k (comme dans le cas de la figure 4.III) ; nous obtenons donc l'inégalité suivante :

$$w_1(G) \leq \sum_{x \in \mathcal{D}} \lceil \log(d(x) + 2) \rceil,$$

où $d(x)$ désigne le degré du sommet x. On a ainsi

$$w_1(G) \leq \gamma_G \cdot \lceil \log(\Delta_{\max} + 2) \rceil,$$

où γ_G et $\Delta_{\max}$ sont respectivement le nombre de domination et le degré maximal de G.

4.3.4 Lien avec les codes superposés

Si $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ est une famille de parties d'un ensemble X, nous disons que $\mathcal{S}$ est une famille ℓ-superposée si, dès que I et J sont deux parties distinctes de $\{1, 2, \cdots, k\}$ de taille au plus ℓ, on a

$$\cup_{i \in I} S_i \neq \cup_{j \in J} S_j.$$

Si $\mathcal{W}$ est un système de $(r, \leq \ell)$-contrôle de G, alors l'ensemble des étiquettes $L_{\mathcal{W}}(x)$ des sommets $x \in V(G)$ est une famille ℓ-superposée. Ces familles ont été introduites dans [66] (sous la forme équivalente de *codes ℓ-superposés*) , et leur lien avec les codes $(r, \leq \ell)$-identifiants a été observé par exemple dans [53].
Le résultat suivant nous sera utile :

Théorème 4.1 (D'yachkov, Rykov [44])**.** *Soit $s_{w,\ell}$ la taille maximale d'une famille ℓ-superposée dans un ensemble de taille w. Il existe deux constantes k_1 et k_2 telles que pour tous w et ℓ strictement positifs on ait*

$$2^{k_1 \frac{1}{\ell^2} w} \leq s_{w,\ell} \leq 2^{k_2 \frac{\log \ell}{\ell^2} w}.$$

4.3.5 Propriétés de monotonie

Les systèmes de contrôle sont beaucoup plus souples que les codes identifiants, et en particulier ils jouissent de propriétés de monotonie. En particulier :

- si $r \leq r'$ et $\ell \geq \ell'$, un système de $(r, \leq \ell)$-contrôle de G est un système de $(r', \leq \ell')$-contrôle de G et donc

$$w_{r,\ell}(G) \geq w_{r',\ell'}(G).$$

- si G est un graphe partiel de G' alors pour tous $r \geq 1$ et $\ell \geq 1$

$$w_{r,\ell}(G') \leq w_{r,\ell}(G).$$

Autrement dit, ajouter une arête à un graphe ne fait que fournir de nouvelles possibilités pour les systèmes de contrôle et fait donc décroître la taille des systèmes minimaux. Ceci rend possible l'étude de graphes *maximaux*, où l'ajout de toute arête fait strictement décroître le cardinal des systèmes de contrôle optimaux ; c'est ce que nous ferons en 4.4.2.

4.3.6 Compression

On dira qu'un système de contrôle $\mathcal{W}$ d'un graphe G est *compressé* si, pour tout sommet $x \in V(G)$ et toute partie non vide $A \subseteq L_{\mathcal{W}}(x)$, il existe un sommet $y \in V(G)$ tel que $L_{\mathcal{W}}(y) = A$.

Étant donné un système de r-contrôle de G, il est toujours possible d'obtenir un système de r-contrôle de G qui soit compressé et de même taille que $\mathcal{W}$: il suffit, dès qu'il existe un sommet $x \in V(G)$ et une partie A non vide telle que $A \subset L_{\mathcal{W}}(x)$ mais ne correspondant à aucune étiquette, de supprimer x des zones de contrôle des contrôleurs de $L_{\mathcal{W}}(x) \setminus A$. On obtient alors un système de contrôle pour lesquels tous les sommets ont même étiquette qu'avec $\mathcal{W}$, sauf x ayant maintenant l'étiquette A. En répétant ce procédé, on obtient en un nombre fini d'étapes un système de contrôle compressé.

Si $\mathcal{W}$ est un système de r-contrôle compressé de G, alors pour tout sommet x on a

$$|L_{\mathcal{W}}(x)| \leq \lfloor \log(|B(x, 2r)| + 1) \rfloor,$$

que l'on déduit immédiatement de la propriété de compression. Un système compressé peut donc être utile pour l'estimation de la taille de certains systèmes de contrôle.

4.3.7 Complexité algorithmique

Sans surprise, la recherche d'un système de $(1, \leq 1)$-contrôle optimal dans un graphe donné est un problème $\mathcal{NP}$-difficile. Mais nous allons un peu plus loin en prouvant un résultat analogue au théorème 3.2 concernant les codes identifiants :

Théorème 4.2 (Auger, Charon, Hudry, Lobstein, annexe F)**.** *Le problème consistant à déterminer, étant donné un entier k et un graphe G planaire et de degré maximal au plus 3, si G admet un système de contrôle de cardinal au plus k, est $\mathcal{NP}$-complet.*

Notons que, pour l'instant, les cas $r > 1$ ou $\ell > 1$ sont ouverts.

4.4 Bornes en fonction du nombre de sommets

4.4.1 Bornes inférieures

Pour $\ell = 1$, la borne du théorème 1.7 s'étend clairement au cas des systèmes de contrôle :

Théorème 4.3. *Soient G un graphe d'ordre n et $r \geq 1$. Alors*

$$\lceil \log(n+1) \rceil \leq w_r(G).$$

L'égalité est atteinte, par exemple, pour tous les graphes de diamètre inférieur ou égal à r. Lorsque $\ell > 1$, nous faisons appel au théorème 4.1 pour obtenir :

Théorème 4.4. *Il existe une constante strictement positive c_1 telle que pour $r > 0$, $\ell > 1$ et tout graphe G d'ordre n on ait*

$$c_1 \frac{\ell^2}{\log \ell} \log n \leq w_{r,\ell}(G).$$

La borne inférieure du théorème 4.1 peut, elle, être utilisée pour construire des systèmes de $(r, \leq \ell)$-contrôle de petite taille dans des graphes particuliers (par exemple des graphes de diamètre inférieur à r).

4.4.2 Bornes supérieures

Un code identifiant définissant de manière naturelle un système de contrôle (voir 4.3.1), le théorème 1.8 prouve que dans tout graphe sans r-jumeaux connexe à n sommets, avec $n \geq 3$, un système de contrôle minimal est au plus de cardinal $n - 1$. On peut prouver en fait bien mieux :

Théorème 4.5 (Auger, Charon, Hudry, Lobstein, annexe F)**.** *Soit G un graphe connexe d'ordre n. Alors*
- *si $n = 1$, $w_1(G) = 1$;*
- *si $n = 2$ ou $n = 3$, $w_1(G) = 2$;*
- *si $n = 4$ ou $n = 5$, $w_1(G) = 3$;*
- *si $n \notin \{1, 2, 4\}$, $w_1(G) \leq \lfloor \frac{2n}{3} \rfloor$.*

Grâce à la propriété de monotonie, il suffit de prouver la borne supérieure de $\frac{2n}{3}$ dans le cas des arbres. D'autre part, la valeur $\lfloor \frac{2n}{3} \rfloor$ est atteinte pour tout $n \notin \{1, 2, 4\}$; nous avons caractérisé, dans [8], les graphes qui l'atteignent. Toujours grâce à la monotonie, il nous suffit de caractériser les graphes G d'ordre n tels que $w_1(G) = \lfloor \frac{2n}{3} \rfloor$ et dont le nombre

d'arêtes est maximal ; tous les graphes satisfaisant exactement cette borne seront alors exactement les graphes partiels connexes de ces graphes maximaux.
Le résultat dépend du reste de n modulo 3. Si ce reste est 0, le résultat est simple à énoncer :

Théorème 4.6 (Auger, Charon, Hudry, Lobstein, [8])**.** *Les graphes connexes G d'ordre n divisible par 3 tels que $w_1(G) = \frac{2n}{3}$ sont précisément les graphes partiels connexes du graphe H construit de la manière suivante : on considère l'union de $\frac{n}{3}$ triangles K_3 disjoints, on choisit un sommet dit point d'attache sur chaque triangle et on relie tous les points d'attache entre eux.*

La figure 4.IV représente le graphe H du théorème ci-dessus, ainsi qu'un exemple de sous-graphe connexe qui atteint la borne de $\frac{2n}{3}$; chaque triangle nécessitant au moins deux contrôleurs, il est immédiat de constater que la borne est atteinte.

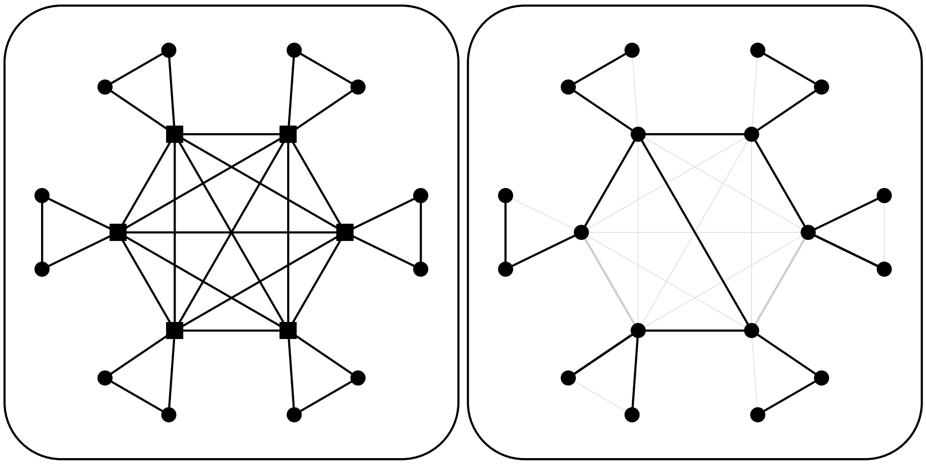


Figure 4.IV – *À gauche, le graphe H d'ordre 18 dont tous les graphes G d'ordre 18 atteignant la borne $w_1(G) = 12$ sont des graphes partiels connexes. Les sommets représentés par des carrés sont les points d'attache des triangles. À droite, un arbre atteignant cette borne.*

Nous appelons *gadgets* les triangles qui composent le graphe H du théorème 4.6. Lorsque n n'est pas divisible par 3, la situation se complique et de nouveaux gadgets apparaissent.

Théorème 4.7 (Auger, Charon, Hudry, Lobstein, annexe F)**.** *Soit G un graphe connexe d'ordre $n = 3k + 2$ où $k \geq 3$. Alors $w_1(G) = 2k + 1$ si et seulement si G est un graphe partiel connexe de l'un des graphes suivants :*

- *la clique d'ordre $k+1$, sur laquelle sont rattachés chacun par un sommet k triangles K_3 ainsi qu'une arête K_2 ;*
- *la clique d'ordre k, sur laquelle sont rattachés chacun par un sommet $k-1$ triangles K_3 ainsi qu'une clique K_5.*

Le premier des graphes ci-dessus est donc construit à partir de $k+2$ gadgets, qui sont des triangles K_3 et une arête K_2. Le deuxième est lui construit à partir de $k+1$ gadgets, à savoir des triangles et une clique K_5. Dans chaque cas chacun des gadgets est muni d'un seul point d'attache, et tous les points d'attaches sont deux à deux adjacents. Les deux graphes extrémaux correspondant au cas $n = 20$ sont représentés sur la figure 4.V.

Dans le cas $n \equiv 1 \mod 3$, de nouveaux gadgets apparaissent, certains disposant de deux points d'attache. Aux arêtes K_2, triangles K_3 et clique K_5 nous devons ajouter :

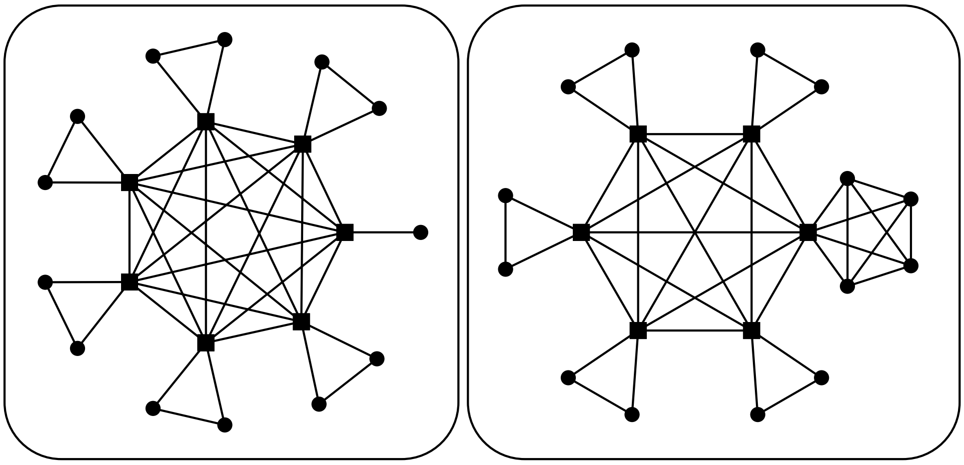


Figure 4.V – *Les deux graphes maximaux d'ordre* 20 *tels que* $w_1 = 13$. *Tout graphe connexe d'ordre* 20 *atteignant la borne du théorème 4.6 est un graphe partiel connexe de l'un de ces deux graphes. Les sommets représentés par des carrés sont les points d'attache des gadgets.*

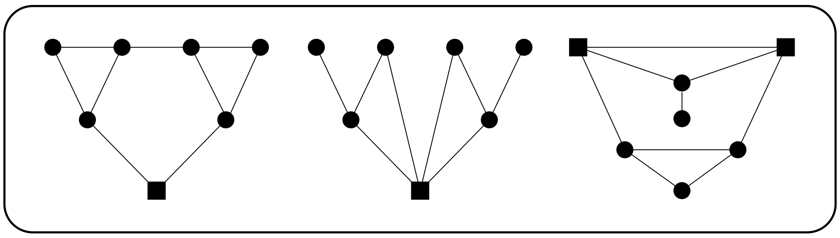


Figure 4.VI – *Les trois gadgets d'ordre* 7. *Les carrés indiquent les points d'attache des gadgets.*

- la clique K_4, qui dispose de deux points d'attache ;
- les trois gadgets d'ordre 7 représentés sur la figure 4.VI. Les deux premiers ont un seul points d'attache et le troisième en a deux.

Le cas $n \equiv 1 \mod 3$ n'est pas terminé à ce jour. Nous caractérisons dans ce cas les arbres atteignant la borne (voir l'annexe D), et donnons des constructions avec ces gadgets que nous pensons suffisantes pour recouvrir tous les cas. Nous conjecturons :

Conjecture 4.8. *Soit G un graphe connexe d'ordre $n = 3k + 1$ où $k \geq 3$. Alors $w_1(G) = 2k + 1$ si et seulement si G est un graphe partiel connexe d'un graphe construit à partir de gadgets pour un des cas suivants :*
- *$k - 1$ triangles K_3 et deux arêtes K_2 ;*
- *$k - 2$ triangles K_3, une arête K_2 et une clique K_5 ;*
- *$k - 3$ triangles K_3 et deux cliques K_5 ;*
- *$k - 1$ triangles K_3 et une clique K_4 (avec deux points d'attache) ;*
- *$k - 2$ triangles K_3 et un gadget d'ordre 7 (fig. 4.VI).*

4.5 Systèmes de contrôle dans les chaînes et les cycles

Pour $r = 1$ et $\ell = 1$, un système de contrôle dans la chaîne $\mathcal{P}_n$ ne peut faire mieux qu'un code identifiant (voir la table 1.III) :

Théorème 4.9 (Auger, Charon, Hudry, Lobstein, annexe F). *Pour tout $n \geq 0$, on a*

$$w_1(\mathcal{P}_n) = \lceil \frac{n+1}{2} \rceil.$$

Pour $\ell = 2$, nous savons que la chaîne $\mathcal{P}_n$ admet des 2-jumeaux puisqu'elle a des sommets de degré 1. En revanche, elle admet comme tout graphe des systèmes de $(1, \leq 2)$-contrôle et nous prouvons :

Théorème 4.10 (Auger, Charon, Hudry, Lobstein, annexe F). *On a $w_{1,2}(\mathcal{P}_n) = n$ pour $1 \leq n \leq 10$ et*

$$w_{1,2}(\mathcal{P}_n) = \lceil \frac{5(n+1)}{6} \rceil$$

pour $n \geq 11$.

La figure 4.VII présente un système de $(1, \leq 2)$-contrôle de cardinal 10 dans une chaîne à 11 sommets, généralisable en un système de taille $5k$ dans une chaîne à $6k - 1$ sommets. Le même type de construction existe pour les chaînes ayant d'autres longueurs.

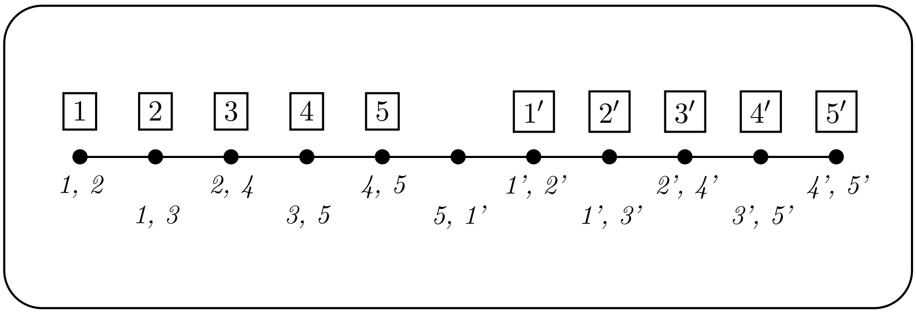


Figure 4.VII – *Un système de $(1, \leq 2)$-contrôle optimal dans les chaînes.*

Pour les valeurs supérieures de ℓ, nous ne pouvons faire mieux qu'utiliser un système de contrôle trivial dans la chaîne :

Théorème 4.11 (Auger, Charon, Hudry, Lobstein, annexe F). *Pour tout $n \geq 0$ et tout $\ell \geq 3$, on a*

$$w_{1,\ell}(\mathcal{P}_n) = n.$$

Nous obtenons des résultats sensiblement équivalent dans les cycles :

Théorème 4.12 (Auger, Charon, Hudry, Lobstein, annexe F). *Pour tout $n \geq 1$ on a*

$$w_1(\mathcal{C}_n) = \lceil \frac{n}{2} \rceil.$$

D'autre part $w_{1,2}(\mathcal{C}_6) = 6$, et pour $n \neq 6$ on a

$$w_{1,2}(\mathcal{C}_n) = w_{1,2}(\mathcal{P}_n) = \lceil \frac{5n}{6} \rceil.$$

4.6 Systèmes de contrôle dans la grille du roi

Les résultats de cette section ont été obtenus en collaboration avec Iiro Honkala lors d'un séjour doctoral à l'université de Turku, Finlande. Ce travail est encore en cours.

Nous renvoyons au chapitre 1 pour la définition de la grille du roi. Ce graphe est infini mais toutes les définitions données pour les sytèmes de contrôles dans les graphes finis s'y généralisent sans problème. Comme pour les codes identifiants, nous nous intéressons à la *densité* des systèmes de contrôle plutôt qu'à leur cardinal (qui sera également infini). La densité d'un système de contrôle $\mathcal{W}$ de la grille du roi est définie de manière analogue à celle d'un code identifiant (voir 1.8).

4.6.1 Identification des sommets

À distance 1, nous pouvons montrer que les codes identifiants sont optimaux comme systèmes de $(1, \leq 1)$-contrôle dans la grille du roi (voir le théorème 1.10). Plus précisément :

Théorème 4.13 (Auger, Honkala)**.** *La densité optimale d'un système de* $(1, \leq 1)$*-contrôle dans la grille du roi est* $\frac{2}{9}$.

La preuve de ce résultat est une adaptation de la preuve du résultat analogue pour les codes identifiants, issue de [40].

Quand r grandit, les systèmes de $(r, \leq 1)$-contrôle deviennent rapidement bien meilleurs que les codes identifants, dont la densité optimale est $\frac{1}{4r}$ pour $r > 1$ (voir la table 1.VI).

Théorème 4.14 (Auger, Honkala)**.** *Soit D_r la densité optimale d'un système de* $(r, \leq 1)$*-contrôle dans la grille du roi. On a pour tout* $r \geq 1$

$$\frac{\log r}{2r^2}(1 + o(1)) \leq D_r \leq \frac{\lceil \log((2r + 1)^2 + 1) \rceil}{(2r + 1)^2}$$

et donc

$$D_r \underset{r\to\infty}{\sim} \frac{\log r}{2r^2}.$$

Preuve. Pour la borne supérieure, nous utilisons le même principe que dans la partie 4.3.3, mais à distance r : l'ensemble

$$\mathscr{C} = (2r + 1)\mathbb{Z} \times (2r + 1)\mathbb{Z}$$

est tel qu'il contient un élément à distance inférieure à r de tout sommet de la grille ; autrement dit, l'union des boules de centre r centrées sur les sommets de $\mathscr{C}$ couvre entièrement $\mathbb{Z}^2$. Or il suffit de $\lceil \log((2r + 1)^2 + 1) \rceil$ contrôleurs au centre de chacune de ces boules pour en identifier les $(2r + 1)^2$ éléments. La densité de $\mathscr{C}$ étant par ailleurs égale à $\frac{1}{(2r+1)^2}$, nous obtenons ainsi le majorant annoncé.

Considérons maintenant $\epsilon > 0$. Pour identifier les sommets d'une boule de centre x fixé et de rayon $\lceil \epsilon r \rceil$, le nombre de contrôleurs nécessaire est au moins égal à $\log((2\epsilon r+1)^2+1)$; d'autre part ces contrôleurs sont nécéssairement centrés dans la boule de centre x et de rayon $\lceil \epsilon r + r \rceil$. En pavant la grille avec des boules de ce rayon on voit donc que

$$\frac{\log((2\epsilon r + 1)^2 + 1)}{(2\lceil (1 + \epsilon)r \rceil + 1)^2} \leq D_r.$$

Ceci étant vrai pour tout $\epsilon > 0$, un peu d'analyse permet d'établir simplement le minorant. □

4.6.2 Identification des ensembles de sommets

Résultats d'impossiblité

Comme tout graphe, la grille du roi admet des systèmes de $(r, \leq \ell)$-contrôle pour toutes les valeurs de r et ℓ ; cependant il est raisonnable de penser que pour ℓ grand, quand r

est fixé, ces systèmes seront tous triviaux, c'est-à-dire de densité supérieure ou égale à 1. Nous avons commencé par le cas $r = 1$, en nous intéressant à la question suivante : *jusqu'à quelle valeur de ℓ la grille du roi admet-elle des systèmes de $(1, \leq \ell)$-contrôle non triviaux ?*

Si $\mathcal{W}$ est un système de contrôle d'un graphe G, nous appelons *ermite* un sommet $h \in V(G)$ tel qu'il existe un contrôleur $w \in \mathcal{W}$ qui contrôle uniquement h. Remarquons qu'il est alors inutile que tout autre contrôleur contrôle h, et qu'on peut ainsi supposer que l'étiquette $L_{\mathcal{W}}(h)$ est un singleton. On peut aussi supposer que le centre de contrôle du contrôleur d'un ermite h est le sommet h lui-même, ce que nous ferons désormais.

Nous avons besoin d'utiliser, pour les propriétés à venir, la notion d' ensemble couvrant une famille de parties. Si X est un ensemble fini et $\mathcal{F}$ est une famille de parties de X, un *ensemble couvrant* $\mathcal{F}$ est un sous-ensemble $\mathscr{C}$ de X tel que, pour tout $F \in \mathcal{F}$, on ait $F \cap \mathscr{C} \neq \emptyset$. Cette notion généralise celle d'ensemble de sommets couvrant un ensemble d'arêtes, ou d'ensemble dominant un graphe.

Supposons r fixé. Soient G un graphe, x un sommet de G et $w_1, w_2, \cdots, w_k$ des r-contrôleurs de G. Nous notons

$$\kappa_x(\{w_1, w_2, \cdots w_k\})$$

le nombre minimal de sommets *distincts de x* nécessaires pour couvrir l'ensemble des parties

$$\{Z(w_i) : 1 \leq i \leq k\}.$$

Clairement, on a $\kappa_x(\{w_1, w_2, \cdots w_k\}) < +\infty$ si et seulement si les contrôleurs w_i, pour $1 \leq i \leq k$, contrôlent x et au moins un autre sommet.
Nous pouvons alors énoncer un lemme qui nous sera utile :

Lemme 4.15. *Soit $\mathcal{W}$ un système de $(r, \leq \ell)$-contrôle dans un graphe G. Alors pour tout $x \in V(G)$, si x n'est pas un ermite, on a*

$$\kappa_x(L_{\mathcal{W}}(x)) \geq \ell$$

et en particulier on a $|B(x, 2r)| \geq \ell + 1$.

Preuve. Si x n'est pas un ermite, alors on a $\kappa_x(L_{\mathcal{W}}(x)) < +\infty$: soit $A \subset V(G) \setminus \{x\}$ couvrant $L_{\mathcal{W}}(x)$ et de taille minimale. Par définition, tout contrôleur $w \in L_{\mathcal{W}}(x)$ contrôle un élément de A ; on a donc $L_{\mathcal{W}}(x) \subset L_{\mathcal{W}}(A)$ et ainsi $L_{\mathcal{W}}(A) = L_{\mathcal{W}}(A \cup \{x\})$; ceci implique que $|A| + 1 > \ell$.

Remarquons maintenant que tous les contrôleurs de x ont leur centre dans $B_G(x, r)$, et donc que leur zone de contrôle est incluse dans $B_G(x, 2r)$. Ainsi $B_G(x, 2r) \setminus \{x\}$ couvre $L_{\mathcal{W}}(x)$ et est donc de cardinal au moins ℓ.

□

Ce lemme implique, par exemple, que dans un graphe de degré maximal Δ, tous les systèmes de $(1, \leq \ell)$-contrôle avec $\ell > \Delta^2$ sont triviaux. Pour la grille du roi, nous obtenons une borne plus fine :

Théorème 4.16 (Auger, Honkala)**.** *Pour $\ell \geq 6$, tous les systèmes de $(1, \leq \ell)$-contrôle dans la grille du roi ont une densité supérieure ou égale à 1.*

Preuve. Soit $\ell \geq 6$ et $\mathcal{W}$ un système de $(1, \leq \ell)$-contrôle dans la grille du roi. Considérons qu'un contrôleur (c, Z) confère un poids de $\frac{1}{|Z|}$ à chaque sommet de Z ; le poids d'un sommet étant alors la somme des poids qui lui sont conférés par chacun des contrôleurs

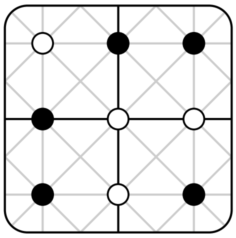


Figure 4.VIII – *Motif M utilisé pour créer un système de $(1, \leq 4)$-contrôle, composé des cinq sommets noirs.*

qui le contrôlent. Il suffit alors, pour prouver qu'un système de contrôle a une densité supérieure ou égale à 1, de montrer que tous les sommets ont un poids supérieur ou égal à 1.

Considérons d'abord un sommet x tel que $|L_{\mathcal{W}}(v)| = 1$; alors, puisque $\ell > 1$ le contrôleur qui contrôle x ne peut contrôler d'autre sommet (ainsi x est un ermite). Ce contrôleur confère donc un poids de 1 à x. D'après le lemme 4.15, si x n'est pas un ermite alors nous devons avoir $|L_{\mathcal{W}}(x)| \geq \ell \geq 6$.
• Supposons que $|L_{\mathcal{W}}(x)| \geq 9$: alors le poids de x est au moins égal à $9 \times \frac{1}{9} = 1$.
• Supposons maintenant $|L_{\mathcal{W}}(x)| = 8$, et posons

$$L_{\mathcal{W}}(x) = \{w_1, w_2, \cdots, w_8\}.$$

Si trois de ces contrôleurs, sans perte de généralité w_1, w_2 et w_3 contrôlent un même sommet distinct de x, alors $\kappa_x(\{w_1, w_2, w_3\}) = 1$ et donc $\kappa_x(\{w_4, w_5, w_6, w_7, w_8\}) \geq 5$. Les cinq contrôleurs w_i pour $4 \leq i \leq 8$ doivent donc avoir des zones de contrôle ne s'intersectant deux à deux qu'en x ; ces cinq zones sont incluses dans la boule $B(x, 2)$, de cardinal 25. Étant disjointes, à part en x, elles confèrent à x un poids minimum de $5 \times \frac{1}{1+\frac{24}{5}} = \frac{25}{29}$. Les contrôleurs w_1, w_2 et w_3, quant à eux, confèrent à x un poids d'au minimum $3 \times \frac{1}{9} = \frac{1}{3}$. Le poids de x est donc au moins $\frac{25}{29} + \frac{1}{3} > 1$.

Si, en dehors de x, trois des huit contrôleurs ne contrôlent jamais un même sommet, alors les éléments de $B(x, 2) \setminus \{x\}$ appartiennent à au plus deux zones de contrôles ; ainsi la taille moyenne de ces zones de contrôle est d'au plus $1 + \frac{48}{8} = 7$, d'où un poids d'au moins $\frac{8}{7}$ pour x.
• Supposons maintenant que $|L_{\mathcal{W}}(x)| = 7$. Parmi ces sept contrôleurs, cinq au moins doivent avoir des zones de contrôle deux à deux disjointes en dehors de x, de façon à avoir $\kappa_x(L_{\mathcal{W}}(x)) \geq 6$. Ces cinq contrôleurs confèrent à x un poids d'au moins $\frac{25}{29}$, et les deux autres un poids d'au moins $\frac{2}{9}$, d'où un poids supérieur à 1 pour x.
• Enfin, si $|L_{\mathcal{W}}(x)| = 6$, les six zones de contrôle doivent être disjointes en dehors de x et ainsi le poids de x est d'au moins $6 \times \frac{1}{1+\frac{24}{6}} = \frac{6}{5} > 1$.

□

Une construction pour $\ell = 4$

Proposons maintenant une construction qui prouve qu'il existe des systèmes de $(1, \leq 4)$-contrôle de densité strictement inférieure à 1 dans la grille du roi. Notre construction utilise le motif M représenté sur la figure 4.VIII. Notons $\{m_1, \cdots, m_5\}$ les sommets de M dans un ordre indifférent ; rappelons ce ces sommets sont des éléments de $\mathbb{Z}^2$. Notre motif M a la propriété suivante :

Propriété 4.17. *Les différences $m_j - m_i$, où les entiers i et j sont tels que $1 \leq i < j \leq 5$, sont deux à deux distinctes.*

Cette propriété nous permet de prouver :

Théorème 4.18. *Si $\mathscr{C}$ est un ensemble de sommets de G tels que deux sommets distincts x et y n'appartenant pas à $\mathscr{C}$ sont au moins à distance 5, alors l'ensemble de contrôleurs*

$$\mathcal{W} = \{(c, c + M) : c \in \mathscr{C}\}$$

est un système de $(1, \leq 4)$-contrôle de G.

Un tel ensemble $\mathscr{C}$ avec une densité optimale est par exemple

$$\mathbb{Z}^2 \setminus (5\mathbb{Z} \times 5\mathbb{Z})$$

qui nous fournit un système de contrôle de densité $\frac{24}{25} < 1$.

Preuve. L'intérêt de la propriété 4.17 est qu'elle assure que si les sommets x et y sont distincts, alors $(x + M) \cap (y + M)$ contient au plus un sommet. Soit $\mathscr{C}$ comme dans l'énoncé ; on remarque de plus que pour tout sommet x, l'ensemble $x - M$ contient au plus un élément de $\mathscr{C}$. Par conséquent, l'étiquette $L_{\mathcal{W}}(x)$ de tout sommet x est de taille au moins 4.

Soient X et Y deux ensembles distincts de sommets de taille au plus 4 et prouvons que $L_{\mathcal{W}}(X) \neq L_{\mathcal{W}}(Y)$; c'est clairement le cas si Y est vide. Supposons donc Y non vide, et aussi dans un premier temps que $|X| < 4$. Soit $y \in Y$; puisque pour tout $x \in X$ on a $|L_{\mathcal{W}}(x) \cap L_{\mathcal{W}}(y)| \leq 1$, on en déduit qu'il existe au moins un sommet de $L_{\mathcal{W}}(y)$ qui n'appartient pas à $L_{\mathcal{W}}(X)$. On a donc $L_{\mathcal{W}}(Y) \neq L_{\mathcal{W}}(X)$.

Nous pouvons ainsi supposer que $|L_{\mathcal{W}}(X)| = |L_{\mathcal{W}}(Y)| = 4$. Il existe alors un sommet $x_1 \in X \setminus Y$ et un sommet $y_1 \in Y \setminus X$. Si jamais nous avions $|L_{\mathcal{W}}(x_1)| = 5$, alors par un argument similaire au précédent nous ne pourrions avoir $L_{\mathcal{W}}(x_1) \subset L_{\mathcal{W}}(Y)$, et inversement si $|L_{\mathcal{W}}(y_1)| = 5$. On peut donc supposer que

$$|L_{\mathcal{W}}(x_1)| = |L_{\mathcal{W}}(y_1)| = 4.$$

Si ces deux étiquettes ne sont pas de cardinal 5, c'est qu'il existe deux sommets de la forme $x_1 + m_1$ et $y_1 + m_2$, où m_1 et m_2 sont dans M, qui n'appartiennent pas à $\mathscr{C}$. Si ces sommets sont distincts, alors $d_G(x_1 + m_1, y_1 + m_2) \geq 5$, et donc $d_G(x_1, y_1) \geq 3$ d'où $L_{\mathcal{W}}(x_1) \cap L_{\mathcal{W}}(y_2) = \emptyset$. Mais alors $L_{\mathcal{W}}(x_1) \subset L_{\mathcal{W}}(Y \setminus \{y_1\})$ et nous pouvons appliquer le même argument que précédemment.

Si jamais $x_1 + m_1 = y_1 + m_2$, alors l'intersection de $x_1 - M$ et $y_1 - M$ est précisément réduite à ce sommet. On a donc $L_{\mathcal{W}}(x_1) \cap L_{\mathcal{W}}(y_1) = \emptyset$, et encore une fois le même argument implique qu'il est impossible d'avoir $L_{\mathcal{W}}(x_1) \subset L_{\mathcal{W}}(Y)$. □

Plus généralement, on peut prouver le résultat suivant, où l'hypothèse de *différences distinctes* est la généralisation de la propriété 4.17.

Théorème 4.19 (Auger, Honkala)**.** *Soit $\ell \geq 2$ et $M = \{m_1, m_2, \cdots, m_{\ell+1}\}$ un sous-ensemble de la boule de rayon r et de centre $\mathbf{0}$ de la grille du roi, ayant des différences distinctes. Alors il existe un système de $(r, \leq \ell)$-contrôle dans la grille de densité*

$$D = 1 - \frac{1}{(4r+1)^2}$$

où toutes les zones de contrôle sont de la forme $x + M$ pour un sommet x.

Le cas $\ell = 5$ est pour l'instant encore ouvert, car nous ne disposons ni d'une construction, ni d'une preuve d'inexistence, d'un système de $(1, \leq 5)$-contrôle non trivial dans la grille du roi.

Deuxième partie

Puissances de graphes

Dans les deux parties précédentes, nous avons utilisé la notion de puissance G^r d'un graphe G afin de nous ramener, de l'étude des codes $(r, \leq \ell)$-identifiants ou des systèmes de $(r, \leq \ell)$-contrôle dans G pour $r > 1$, à l'étude des codes $(1, \leq \ell)$-identifiants ou des systèmes de $(1, \leq \ell)$-contrôle dans G^r. Les puissances de graphes nous permettent donc d'intégrer les difficultés liées aux calculs de distances au graphe considéré et de se focaliser sur l'étude des codes ou systèmes. Ils permettent ainsi de transposer de nombreux résultats, établis dans le cas $r = 1$, au cas général où r est quelconque, simplement en les appliquant à G^r au lieu de G.

Au-delà de cet aspect pratique, les graphes qui sont des puissances jouissent de certaines propriétés combinatoires intéressantes. Citons par exemple le résultat suivant, dû à H. Fleischner, dont il existe plusieurs variantes :

Théorème 4.20 (Fleischner [47])**.** *Le carré d'un graphe* 2*-connexe admet un cycle hamiltonien.*

Nous présenterons dans cette partie nos différents résultats concernant les puissances de graphes. Le chapitre 5 expose des résultats de type extrémal : nous étudions quelles configurations présentent un nombre d'arêtes maximum ou minimum pour certaines propriétés liées aux puissances. Le chapitre 6 présentera des résultats algorithmiques et combinatoires liés plus particulièrement à la notion de carré d'un graphe.

Rappelons la définition de la puissance r-ième d'un graphe non orienté G adoptée ici : c'est le graphe G^r, dont l'ensemble des sommets est le même que celui de G et tel que deux sommets distincts x et y sont reliés par une arête dans G^r si et seulement si leur distance $d_G(x, y)$ dans G vérifie

$$d_G(x, y) \leq r.$$

Notons que le diamètre de G peut être interprété comme le plus petit entier r tel que toutes les composantes connexes de G^r soient des graphes complets. A titre d'exemple, la figure 4.IX présente les différentes puissances d'un graphe.

Remarquons que $G^1 = G$ et qu'on dispose des relations habituelles justifiant la notation exponentielle : $(G^s)^t = G^{st}$ pour tous les entiers naturels non nuls s et t. Puisque $V(G) = V(G^r)$, les arêtes de G sont des arêtes de G^r lorsque $r \geq 1$: ainsi G est un graphe partiel de G^r. Plus généralement, G^s est un sous-graphe partiel couvrant de G^t si $s \leq t$.

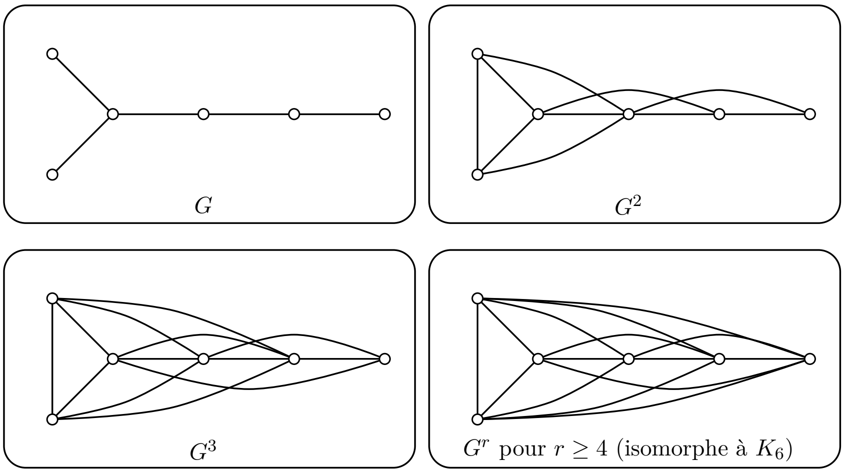


Figure 4.IX – *Un graphe G et ses puissances G^2, G^3 et G^r pour $r \geq 4$.*

Si H est un graphe tel que $H^r = G$, on dit que H est une racine r-ième de G ; dans le cas $r = 2$, on parle de carré et de racines carrées. Deux racines r-ièmes distinctes

H_1 et H_2 d'un graphe G (pour la même valeur de r) seront appelées *coracines*. Deux coracines étant définies sur le même ensemble de sommets que G, on pourra par exemple considérer des opérations ensemblistes telles que $E(H_1) \cap E(H_2)$. Dans toute cette partie, et particulièrement dans le chapitre 6, il faut garder à l'esprit que l'on travaille sur un ensemble de sommets fixé et pas à isomorphisme près : deux coracines H_1 et H_2 d'un graphe G, même si elles sont isomorphes, seront considérées comme des racines distinctes dès lors que $E(H_1) \neq E(H_2)$. Par exemple, la figure 4.X représente le graphe hyperoctaédral d'ordre 6, qui admet quatre racines carrées, toutes isomorphes à un cycle $\mathcal{C}_6$.

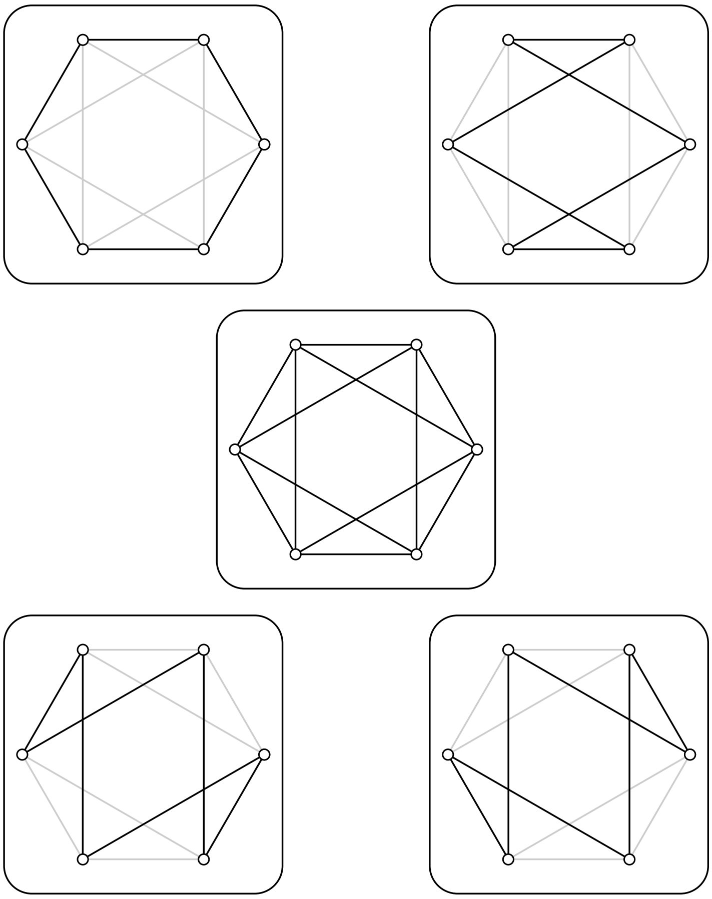


Figure 4.X – *Le graphe hyperocatédral d'ordre* 6 *et ses* 4 *racines carrées isomorphes.*

Chapitre 5

Problèmes extrémaux

Sommaire

5.1 Introduction 53
5.2 État des lieux 53
5.3 Nouveaux résultats dans le cas non orienté 55
5.4 Nouveaux résultats dans le cas orienté 56

5.1 Introduction

Considérons un réseau social, comme les réseaux *Facebook*, *Twitter* ou *MSN*. Supposons qu'à un instant donné, chaque personne ajoute comme amis les amis de ses amis : alors un certain nombre de nouvelles amitiés (connexions) vont se créer dans le réseau. Peut-on évaluer le nombre minimal de connexions qui seront créées, et le nombre minimal d'amitiés qui existeront au final ? Quels sont les cas extrémaux ? On peut aussi répéter l'opération plusieurs fois avant d'envisager ces questions.

Ces différents problèmes se formulent naturellement en termes de puissances de graphes. Voici précisément ce que nous nous proposons d'étudier :

Questions

(I) Quelle est la taille maximale d'un graphe G d'ordre n de diamètre d ?

(II) Quelle est la taille minimale de la puissance r-ième d'un graphe G connexe d'ordre n ?

(III) Étant donné un graphe connexe G d'ordre n et de diamètre $d > r$, quelle est la valeur minimale de

$$|E(G^r)| - |E(G)|,$$

autrement dit combien d'arêtes sont ajoutées au minimum quand on passe de G à sa puissance r-ième ?

Dans chaque cas, nous cherchons la liste exhaustive des graphes réalisant l'extremum. Nous envisagerons ensuite, dans la section 5.4, ces questions dans le cas des graphes orientés.

5.2 État des lieux

Un théorème d'Ore donne la réponse à la question (I) :

Théorème 5.1 (Ore [79]). *Soit G un graphe d'ordre n et de diamètre $d \geq 2$. Alors la taille de G vérifie*

$$|E(G)| \leq \frac{n(n-1)}{2} - b(n, d-1) - 1,$$

où

$$b(n,r) = (r-1)(n-1-\frac{r}{2}).$$

Autrement dit, quel que soit $r \geq 1$, si le nombre d'arêtes qui manquent à un graphe d'ordre n pour être complet est inférieur ou égal à $b(n,r)$, alors son diamètre est au plus r, i.e. G^r est complet. Ore caractérise aussi les graphes réalisant l'égalité dans le théorème 5.1, c'est-à-dire les graphes de diamètre fixé ayant le plus d'arêtes possible. Nous retrouvons ce résultat comme conséquence de notre réponse à la question (III), mais il peut être intéressant de considérer dès maintenant ces graphes extrémaux. Pour cela, nous appelons *chaîne de cliques de longueur q* et notons $(W_0, W_1, \cdots, W_q)$ un graphe G (voir fig. 5.I) tel que :

- les ensembles $W_0, W_1, \cdots, W_q$ forment une partition de $V(G)$;
- pour tout i vérifiant $0 \leq i \leq q-1$ le graphe $G[W_i \cup W_{i+1}]$ est une clique ;
- si $i+1 < j$, il n'y a aucune arête entre W_i et W_j.

On peut construire toute chaîne de cliques de longueur q en partant d'une chaîne de longueur q et en remplaçant chaque sommet par une clique, reliant entre eux tous les sommets de deux cliques correspondant à des sommets initialement adjacents. Le diamètre d'une telle chaîne de clique est égal à q.

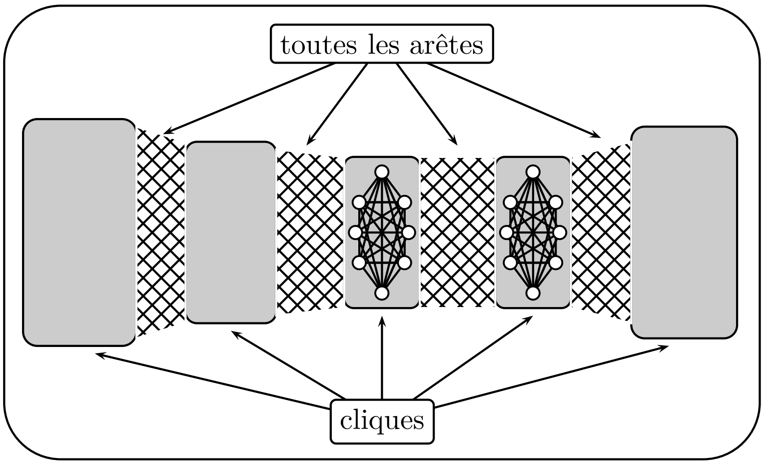


Figure 5.I – *Une chaîne de cliques, cas non orienté*

Avec ce vocabulaire, les graphes extrémaux correspondant à la question (I) sont aisément décrits. Il convient de distinguer les graphes *diamètre-critiques*, pour lesquels l'ajout de toute arête diminuerait strictement le diamètre, et les graphes *maximum diamètre-critiques*, qui de plus ont le nombre maximum d'arêtes possible (et qui sont précisément ceux qui nous intéressent dans le cadre de la question (I)) :

Théorème 5.2 (Ore [79]). *Pour tout $d \geq 2$, un graphe de diamètre d est diamètre-critique si et seulement s'il est une chaîne de cliques de longueur d, et est maximum diamètre-critique si et seulement s'il est une chaîne de cliques de longueur d où toutes les cliques sont réduites à un seul sommet, sauf éventuellement deux cliques consécutives ne pouvant se trouver aux extrémités de la chaîne de cliques.*

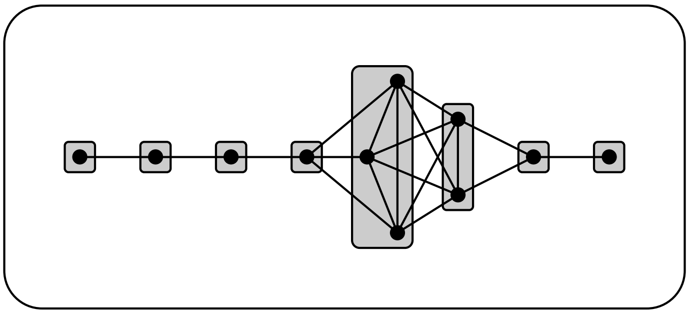


Figure 5.II – *Un graphe maximum diamètre-critique à* 11 *sommets et de diamètre* $d = 7$.

La figure 5.II illustre ce théorème en présentant un exemple de graphe maximum diamètre-critique de diamètre 7 à 11 sommets. On pourra vérifier que ce graphe comporte 19 arêtes, et que la borne du théorème 5.2 vaut

$$\frac{11 \times 10}{2} - 5 \times (11 - 1 - \frac{6}{2}) - 1 = 19.$$

Passons maintenant à la question (II). La réponse à cette question est également connue, mais nous prouverons que les chaînes sont les seuls graphes réalisant l'égalité. Le théorème suivant est cité dans le livre de F. Buckley et F. Harary ([22]) comme corollaire d'un résultat non publié de M.O. Albertson, D. Berman et F. Buckley. Le cas $r = 2$ y est attribué à M. Capobianco.

Théorème 5.3 (Albertson, Berman, Buckley, Capobianco, Harary [22]). *Soit* $r \geq 1$ *et* G *un graphe connexe d'ordre* n. *Alors*

$$|E(G^r)| \geq nr - \frac{r(r+1)}{2}.$$

La question (III), quant à elle, n'a été abordé que dans le cas particulier $r = 2$:

Théorème 5.4 (Aingworth, Motwani, Harary [1]). *Si* G *est un graphe d'ordre* n *et de diamètre au moins* 3, *alors*

$$\left|E(G^2)\right| - |E(G)| \geq n - 2.$$

Les auteurs prouvent que l'égalité est possible, mais ne caractérisent pas les graphes réalisant le maximum, ce que nous ferons.

5.3 Nouveaux résultats dans le cas non orienté

La question (I), comme précisé ci-avant, a été entièrement résolue par Ore ([79]). Pour la question (II), nous avons caractérisé les graphes réalisant l'extremum, complétant ainsi le théorème 5.3 :

Théorème 5.5 (Auger, Charon, Hudry, Lobstein [11] & appendice H). *Soit* $r \geq 1$ *et* G *un graphe connexe d'ordre* $n \geq 1$ *et de diamètre* $d < r$. *Alors*

$$|E(G^r)| = nr - \frac{r(r+1)}{2}$$

si et seulement si G *est une chaîne.*

Ce résultat est aussi une conséquence du théorème 5.6, qui répond à la question (III). Le voici :

Théorème 5.6 (Auger, Charon, Hudry, Lobstein [11] & appendice H)**.** *Soit $r \geq 1$. Si G est un graphe connexe d'ordre n et de diamètre au moins $r+1$, alors*

$$|E(G^r)| - |E(G)| \geq b(n,r),$$

où

$$b(n,r) = (r-1)(n-1-\frac{r}{2}).$$

Si G^r n'est pas complet alors G gagne donc au moins $b(n,r)$ arêtes lorsqu'il est élevé à la puissance r, et donc le nombre d'arêtes qui manquent à G pour être complet est strictement supérieur à $b(n,r)$. On retrouve ainsi le théorème 5.1.

Décrivons maintenant les graphes pour lesquels il y a égalité, en termes de chaînes de cliques (cf. section 5.2).

Théorème 5.7 (Auger, Charon, Hudry, Lobstein [11] & appendice H)**.** *Les seuls graphes réalisant l'égalité dans le théorème 5.6 sont les chaînes de cliques $(W_0, W_1, \cdots, W_q)$ appartenant à l'un des types suivants :*

- *type 1 : $q = r+1$ et il existe au plus deux entiers j tels que $|W_j| > 1$, qui doivent alors être consécutifs.*
- *type 2 : $q \geq r+1$ et $|W_i| = 1$ dès que $2 \leq i \leq q-2$; de plus si $q = r+1$ alors $|W_1| = 1$ et $|W_{q-1}| = 1$, et si $q = r+2$ alors $|W_1| = 1$ ou $|W_{q-1}| = 1$.*

Notons que le théorème 5.5 est une conséquence immédiate des théorèmes 5.6 et 5.7, car on vérifie que

$$n - 1 + b(n,r) = nr - \frac{r(r+1)}{2}.$$

Ainsi, un graphe réalisant l'extremum du théorème 5.5 est nécessairement une chaîne de cliques ayant exactement $n-1$ arêtes, et ne peut donc être qu'une chaîne.

Les graphes décrits dans le théorème 5.7 sont aisément compris avec une description conditionnelle : ce sont les chaînes de cliques $(W_0, W_1, \cdots, W_q)$ avec $q \geq r+1$ telles qu'un sommet d'une clique W_i vérifiant $|W_i| \geq 2$ admette dans le graphe exactement un sommet à distance k pour tout k vérifiant $2 \leq k \leq r$.

Parmi les graphes de type 1, on retrouve les graphes maximum diamètre-critiques évoqués dans la section précédente (théorème 5.2, voir la fig. 5.II pour un exemple) puisqu'il leur manque exactement $b(n,r)+1$ arêtes pour êtres complets et qu'il ne manque qu'une seule arête à leur puissance r-ième pour être un graphe complet. À ces graphes s'ajoutent ceux de type 1 ayant une clique comportant plus d'un sommet sur une clique extrême (voir un exemple sur la figure 5.III), ainsi que ceux de type 2. Les différentes possibilités pour les graphes de type 2 sont illustrées sur la figure 5.IV. Notons que les deux types ne sont pas exclusifs.

5.4 Nouveaux résultats dans le cas orienté

Dans le cas orienté, nous avons établi des résultats similaires. Les preuves sont cependant plus complexes ; en effet, les notions de connexité dans un graphe orienté sont plus subtiles que dans le cas non orienté. En particulier, l'hypothèse qui nous permet ici de généraliser nos résultats est celle de forte connexité.

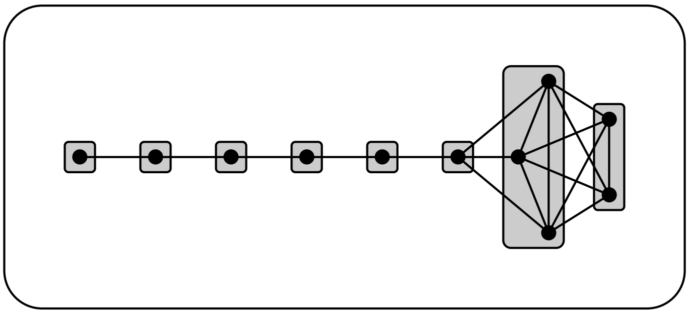


Figure 5.III – *Un graphe de type 1 à 11 sommets pour $r = 6$ (diamètre $r + 1$), qui n'est pas maximum diamètre-critique.*

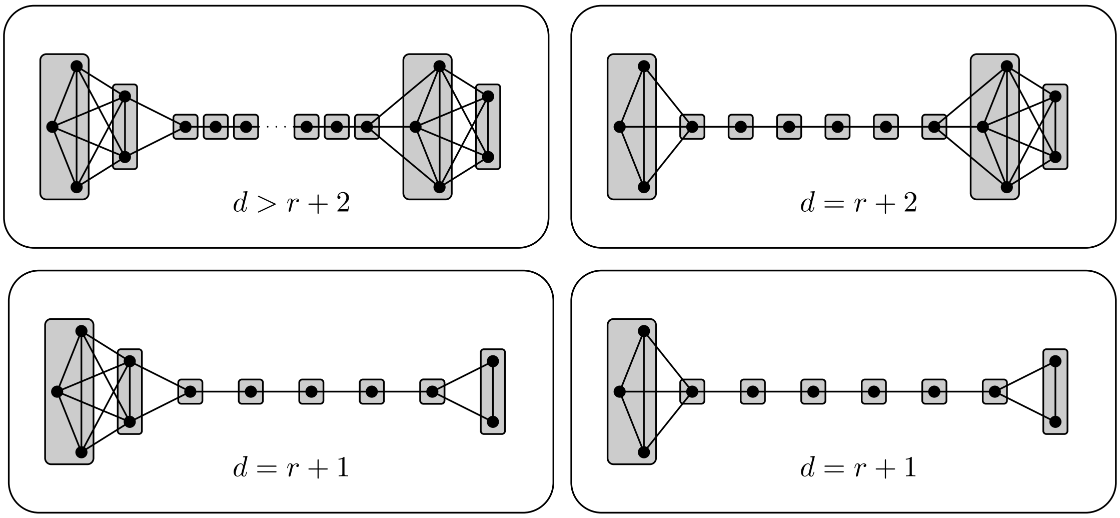


Figure 5.IV – *Différents graphes de type 2, dans le cas $r = 6$.*

Nous appelons *clique orientée* dans un graphe orienté D un sous-ensemble de sommets K de D tel que, pour tous x et y distincts dans K, les arcs (x, y) et (y, x) soient dans $A(D)$. Si $V(D)$ est une clique orientée, on dira que D est un *graphe orienté complet*.

Nous adaptons la notion de chaîne de cliques du cas non orienté (voir section 5.3) en la notion de *chemin de cliques* : un chemin de cliques de longueur q également noté $(W_0, W_1, \cdots, W_q)$, est un graphe orienté D (voir fig. 5.V) tel que :

- les ensembles $W_0, W_1, \cdots, W_q$ forment une partition de $V(D)$;
- pour tout i vérifiant $0 \leq i \leq q - 1$, si $x \in W_i$ et $y \in W_{i+1}$ alors $(x, y) \in A(D)$;
- pour tous les entiers i et j vérifiant $0 \leq i \leq j \leq q$, si $x \in W_i$ et $y \in W_j$ sont distincts alors $(y, x) \in A(D)$;
- il n'y a aucun arc de W_i vers W_j si $j > i + 1$.

Un chemin de cliques est donc une suite de cliques orientées, telle que tous les sommets d'une clique soient reliés aux sommets de la suivante, et telle qu'on ait tous les *arcs retour*, c'est-à-dire tous les arcs des sommets d'une clique W_j vers les sommets des cliques W_i lorsque $j > i$. Notons qu'un chemin de cliques est bien évidemment un graphe orienté fortement connexe.

L'analogue du théorème 5.5 dans le cas orienté est dû à A. Ghouila-Houri. Nous caractérisons dans l'article [10] (appendice I) les graphes atteignant l'extremum.

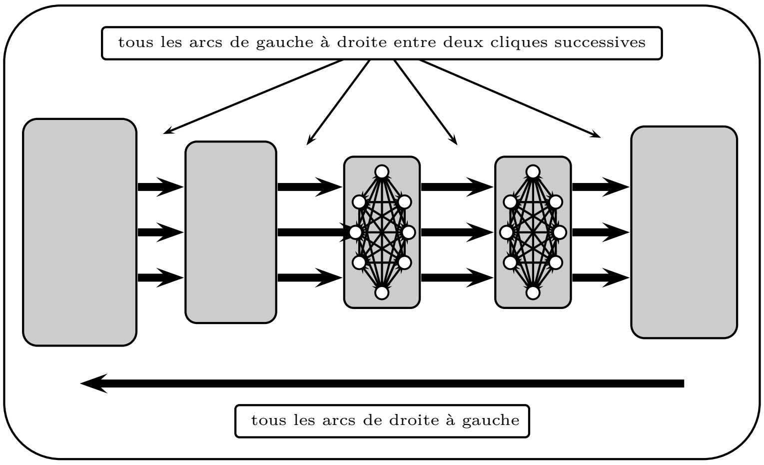


Figure 5.V – *Un chemin de cliques*

Théorème 5.8 (Ghouila-Houri [51]). *Soit D un graphe orienté fortement connexe d'ordre n et de taille m. Alors diam$(D) \leq n-1$ si $n \leq m \leq \frac{n^2+n-2}{2}$ et sinon*

$$diam(D) \leq \lfloor n + \frac{1}{2} - \sqrt{2m - n^2 - n + \frac{17}{4}} \rfloor.$$

Nous prouvons aussi l'analogue des théorèmes 5.6 et 5.7 dans le cas orienté. On a premièrement la borne :

Théorème 5.9 (Auger, Charon, Hudry, Lobstein [10] & appendice I). *Soit $r \geq 1$. Si D est un graphe orienté fortement connexe, d'ordre n et de diamètre au moins $d > r$, alors*

$$|A(D^r)| - |A(G)| \geq b(n,r),$$

où

$$b(n,r) = (r-1)(n-1-\frac{r}{2}).$$

Voici les graphes extrémaux (plus simplement décrits à l'aide du cas non orienté) :

Théorème 5.10 (Auger, Charon, Hudry, Lobstein [10] & appendice I). *Les graphes orientés réalisant l'extremum dans le théorème 5.9 sont obtenus en prenant une chaîne de cliques $(W_0, W_1, \cdots, W_k)$ satisfaisant les conditions du théorème 5.7 (pour la même valeur de r) et en considérant le chemin de cliques correspondant $(W_0, W_1, \cdots, W_k)$.*

Chapitre 6

Racines carrées de graphes

Sommaire

6.1 Introduction et état des lieux 59
6.2 Résultats algorithmiques 60
6.3 Multiplicité des racines carrées 61
6.3.1 Unicité des racines de maille au moins 7 61
6.3.2 Deux familles de graphes 61
6.3.3 Décomposition et borne sur la multiplicité 64

6.1 Introduction et état des lieux

Nous revenons au graphes non orientés dans ce chapitre. La notion de code r-identifiant nous ayant naturellement amenés à celle de puissances de graphes et à l'étude de résultats extrémaux les concernant (chapitre 5), on ne peut maintenant s'empêcher de se poser les questions suivantes :

(I) Un graphe peut-il admettre plusieurs racines r-ièmes distinctes ? Quel est leur nombre ? Quelles sont alors les relations entre ces racines ?

(II) Existe-t-il un algorithme efficace permettant de déterminer si un graphe donné G est une puissance d'un autre graphe ?

Plus généralement, on peut se poser les mêmes questions quand on impose aux racines d'appartenir à une classe de graphes fixée. Le cas $r = 2$ est déjà non trivial. Présentons quelques-uns des résultats déjà connus sur les carrés et racines carrées de graphes lorsque nous avons commencé cette étude. Un des premiers résultats publiés, à notre connaissance, est le suivant :

Théorème 6.1 (Ross, Harary [85])**.** *Un graphe admet, à isomorphisme près, au plus une racine carrée dans la classe des arbres.*

Nous allons étendre ce résultat en montrant qu'en fait, à quelques exceptions près, un graphe admet au plus une racine carrée dans une classe contenant strictement celle des arbres, et ce *en tant que graphe partiel*, ce qui est plus restrictif qu'à isomorphisme près (théorème 6.5). D'autre part, il existe ([72]) un algorithme de complexité linéaire dû à Lin et Skienna permettant de savoir si un graphe est le carré d'un arbre et de déterminer, le cas échant, cette racine carrée. Dans le cas général, les carrés de graphes ont été caractérisés :

Théorème 6.2 (Mukhopadhyay [78]). *Un graphe G d'ordre n, avec*

$$V(G) = \{v_1, v_2, \cdots, v_n\}$$

est un carré si et seulement s'il existe des ensembles de sommets

$$K_1, K_2, \cdots, K_n$$

vérifiant :

1. *quel que soit $1 \leq i \leq n$, on a $v_i \in K_i$;*
2. *quel que soit $1 \leq i \leq n$, le graphe $G[K_i]$ est une clique ;*
3. *quels que soient $1 \leq i \leq n$ et $1 \leq j \leq n$, on a $v_i \in K_j$ si et seulement si $v_j \in K_i$;*
4. *enfin $E(G) = \cup_{i=1}^{n} E(K_i)$.*

Malheureusement, on peut aisément se rendre compte que cette caractérisation ne conduit pas à un algorithme efficace pour reconnaître les carrés de graphes. Et en effet, on a le résultat suivant :

Théorème 6.3 (Motwani, Sudan [77]). *Le problème de décision consistant à déterminer si un graphe donné est un carré est $\mathcal{NP}$-complet.*

La reconnaissance des carrés de graphes quelconques est donc difficile, alors que la reconnaissance des carrés d'arbres peut se faire en temps linéaire. Il existe aussi un résultat de Lau s'intéressant à la classe des graphes bipartis (qui étend la classe des arbres), stipulant qu'on peut reconnaître les carrés de graphes bipartis en temps polynomial ([71]).

6.2 Résultats algorithmiques

Une autre façon d'étendre les résultats connus pour la reconnaissance des carrés d'arbres est de considérer la classe des graphes ayant une maille au moins k, l'entier k étant fixé (les arbres pouvant être considérés comme ayant une maille infinie, c'est-à-dire n'ayant aucun cycle). Nous noterons $\mathcal{G}_k$ cette classe. Les valeurs qui vont particulièrement nous intéresser sont $k = 6$ et $k = 7$. Nous allons voir que les résultats connus pour la classe des arbres s'étendent pour la classe $\mathcal{G}_7$; pour $\mathcal{G}_6$, les choses se compliquent nettement.

Nous renvoyons à l'article ([12] , appendice J) pour les détails, en nous contentant ici de citer le principal résultat de nature algorithmique :

Théorème 6.4 (Auger, Charon, Hudry, Lobstein [12] & appendice J). *Il existe un algorithme polynomial qui détermine si un graphe donné admet des racines carrées dans $\mathcal{G}_6$ et, le cas échant, calcule la liste de ses racines. Cet algorithme fonctionne en temps $O(n^2\Delta_{\min}^2)$ dans le pire des cas, où n et $\Delta_{\min}$ sont respectivement l'ordre et le degré minimal du graphe constituant l'instance.*

Cet algorithme peut être amélioré pour être plus rapide dans $\mathcal{G}_7$. En revanche, la reconnaissance des carrés de graphes de $\mathcal{G}_5$ reste un problème ouvert ; nous conjecturons que ce problème est $\mathcal{NP}$-complet.

6.3 Multiplicité des racines carrées

6.3.1 Unicité des racines de maille au moins 7

Le théorème 6.1 stipule que les racines carrées d'un graphe qui appartiennent à la classe des arbres sont toutes isomorphes. Nous avons étendu ce résultat :

Théorème 6.5 (Auger, Charon, Hudry, Lobstein [12] & appendice J)**.** *Les seuls graphes admettant strictement plus d'une racine carrée dans la classe $\mathcal{G}_7$ (en tant que graphes partiels) sont :*

- *les graphes complets ;*
- *les graphes formés en considérant deux graphes complets de taille au moins 3 partageant une unique arête (voir fig. 6.I).*

De plus, les racines carrées dans $\mathcal{G}_7$ de ces graphes sont isomorphes.

On peut facilement se rendre compte que les racines carrées du graphe complet K_n appartenant à $\mathcal{G}_7$ sont les n étoiles centrées sur les différents sommets du graphe. Quant au deuxième type de graphe exceptionnel, ses racines carrées appartenant à $\mathcal{G}_7$ sont au nombre de deux, du type de celles qui sont représentées sur la figure 6.I. Notons que dans chaque cas, les racines carrées sont des arbres. Les autres graphes, en particulier tous les graphes de rayon au moins 2, admettent au plus une racine carrée dans $\mathcal{G}_7$.

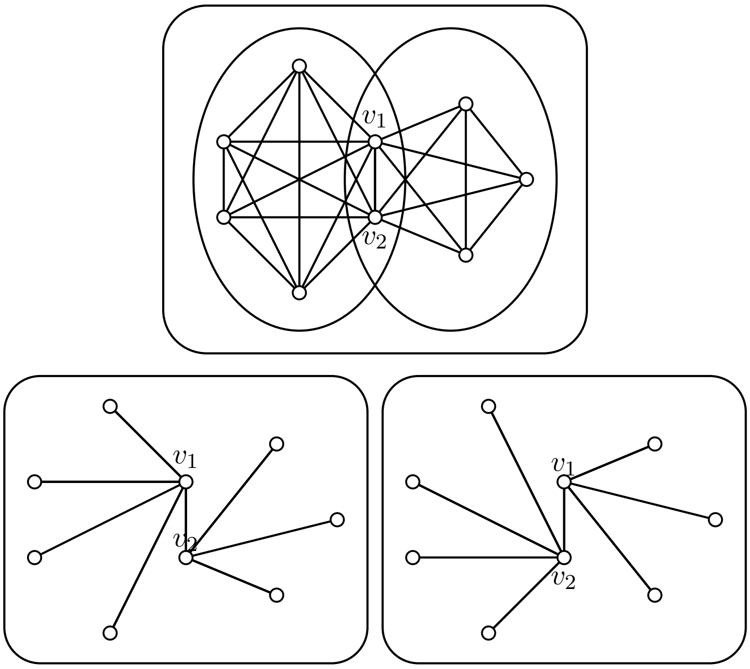


Figure 6.I – *Le deuxième type de graphe exceptionnel du théorème 6.5, ainsi que ses deux racines carrées dans $\mathcal{G}_7$*

6.3.2 Deux familles de graphes

Si les racines carrées dans $\mathcal{G}_7$ sont en général uniques, à l'opposé un graphe peut avoir de nombreuses racines dans $\mathcal{G}_6$, qui seront cependant isomorphes. Dans ce paragraphe nous allons présenter de tels graphes. L'exemple fondamental est celui du graphe hyperoctaédral d'ordre 6, que l'on préférera ici voir comme le carré d'un cycle : celui-ci admet exactement

4 racines carrées distinctes dans $\mathcal{G}_6$, toutes isomorphes à un cycle de longueur 6 (voir la figure 4.X).

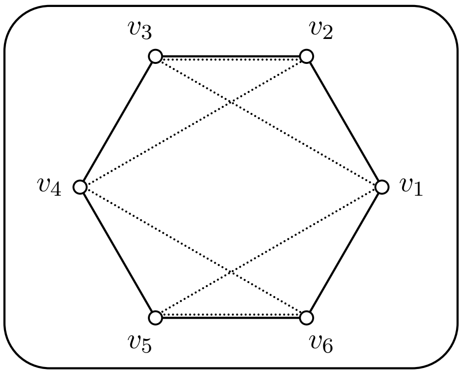


Figure 6.II – *un cycle* $v_1v_2v_3v_4v_5v_6v_1$ *de longueur* 6 *et une de ses coracines* $v_1v_3v_2v_4v_6v_5v_1$.

Sur la figure 6.II, nous avons superposé deux coracines de ce graphe. On peut remarquer qu'un isomorphisme simple échange les deux coracines : c'est l'application qui échange v_2 avec v_3 et v_5 avec v_6, tout en laissant stables v_1 et v_4. Remarquons également que les deux paires de sommets échangés par cet isomorphisme correspondent aux deux arêtes v_2v_3 et v_5v_6 qui sont les seules à appartenir aux deux cycles. Cette situation est en fait tout à fait générale. Nous avons prouvé :

Théorème 6.6 (Auger, Charon, Hudry, Lobstein [12] & appendice J)**.** *Soit G un graphe admettant deux racines carrées distinctes H_1 et H_2 dans $\mathcal{G}_6$. Alors les arêtes appartenant à $E(H_1) \cap E(H_2)$ forment un couplage dans G et l'application qui échange deux à deux les sommets des arêtes de ce couplage et laisse stables les autres sommets est un isomorphisme de H_1 sur H_2.*

En particulier, si un graphe admet plusieurs racines dans $\mathcal{G}_6$, elles sont toutes isomorphes. En fait, on peut aller plus loin et caractériser exactement ce type de couplage dans les racines. Précisément :

Théorème 6.7 (Auger, Charon, Hudry, Lobstein [12] & appendice J)**.** *Soit G un graphe admettant une racine carrée $H_1 \in \mathcal{G}_6$ et soit $D \subset E(G)$ un ensemble d'arêtes de G. Alors il existe une racine carrée $H_2 \in \mathcal{G}_6$, distincte de H_1 et telle que $D = E(H_1) \cap E(H_2)$ si et seulement si D vérifie les conditions suivantes :*

(i) *D est un couplage de G ;*

(ii) *si $e \in E(H_1) \setminus D$, alors e est adajcente à une unique arête appartenant à D ;*

(iii) *si $abca'b'$ est une chaîne dans H_1 avec $ab \in D$ et $a'b' \in D$, alors il existe un sommet c' tel que $abca'b'c'a$ forme un cycle de longueur 6 dans H_1.*

Ces conditions donnent aux graphes admettant plusieurs racines carrées distinctes dans $\mathcal{G}_6$ une structure très particulière. Nous présentons un exemple, plus complexe qu'un simple cycle, sur la figure 6.III. Il s'agit d'un graphe appartenant à $\mathcal{G}_6$, ainsi qu'une de ses coracines.

Ce graphe admet en fait 4 coracines (toujours dans $\mathcal{G}_6$), ce qui porte à 5 le nombre de racines dans $\mathcal{G}_6$ de son carré. Ce type de graphe peut en fait être généralisé de la façon suivante : considérons, pour un choix de $a \geq 2$ et $b \geq 2$ les ensembles $\mathcal{A}$ et $\mathcal{B}$ des parties ayant respectivement a et b éléments parmi les $a + b - 1$ éléments

$$1, 2, \cdots, a + b - 1$$

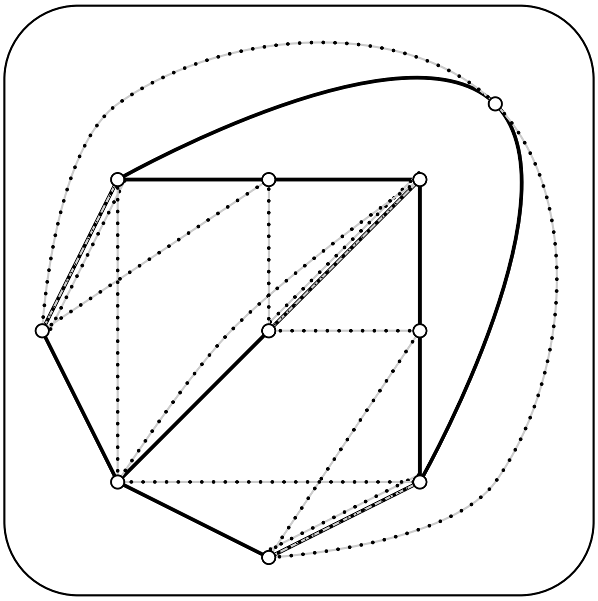


Figure 6.III – *Deux coracines (l'une en noir et l'autre en pointillés). On passe d'un graphe à l'autre par échange des sommets sur chacune de leurs trois arêtes communes.*

(si $a = b$, les ensembles $\mathcal{A}$ et $\mathcal{B}$ sont deux copies disjointes du même ensemble). Étant donnés $A \in \mathcal{A}$ et $B \in \mathcal{B}$, les ensembles A et B ont au moins un élément en commun. On définit alors un graphe $\mathcal{R}_{a,b}$ sur l'ensemble de sommets $\mathcal{A} \cup \mathcal{B}$, en le munissant des arêtes de la forme AB pour tous $A \in \mathcal{A}$ et $B \in \mathcal{B}$ tels que $|A \cap B| = 1$. Les deux graphes de la figure 6.III sont en fait isomorphes à $\mathcal{R}_{2,3}$ comme on peut s'en apercevoir sur la figure 6.IV (et le cycle de longueur 6 était isomorphe à $\mathcal{R}_{2,2}$). En considérant l'ensemble des arêtes de la forme AB avec $A \cap B = \{i\}$, où i est fixé, on obtient un couplage correspondant à une coracine de ce graphe.

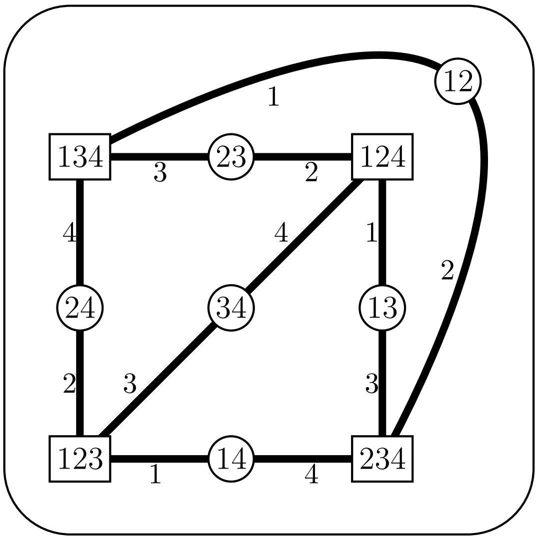


Figure 6.IV – *Le graphe $\mathcal{R}_{2,3}$. Sur les arêtes sont indiquées les étiquettes, correspondant à l'intersection des deux extrémités. L'ensemble des arêtes ayant une étiquette donnée forme un couplage correspondant à une coracine de $\mathcal{R}_{2,3}$.*

On montre, plus généralement :

Théorème 6.8 (Auger, Charon, Hudry, Lobstein [12] & appendice J). *Pour tous $a \geq 2$ et $b \geq 2$, le graphe $\mathcal{R}_{a,b}$ admet exactement $a+b$ racines carrées dans $\mathcal{G}_6$.*

Il existe une autre famille de graphes, les graphes $\mathcal{R}_a$ pour $a \geq 3$, qui sont définis d'une façon analogue et jouissent de propriétés similaires.

6.3.3 Décomposition et borne sur la multiplicité

On peut montrer, réciproquement, que les deux familles de graphes que nous avons définies, les familles $\{\mathcal{R}_{a,b}\}_{a,b\geq 2}$ et $\{\mathcal{R}_a\}_{a\geq 3}$, sont les briques de base permettant de construire des graphes de $\mathcal{G}_6$ admettant de nombreuses coracines. Voici précisément le résultat :

Théorème 6.9 (Auger, Charon, Hudry, Lobstein [12] & appendice J). *Soit $H_0 \in \mathcal{G}_6$ un graphe connexe admettant exactemement r coracines $H_1, H_2, \cdots H_r$ dans $\mathcal{G}_6$, où $r \geq 3$. Soit*

$$D = E(H_0) \cap \left\{ \bigcup_{i=1}^{r} E(H_i) \right\}.$$

Alors les composantes connexes du graphe $(V(H_0), D)$ sont :

- *soit des sommets isolés ;*
- *soit des étoiles à $r+1$ sommets ;*
- *soit isomorphes à $\mathcal{R}_{a,r+1-a}$ pour un a tel que $2 \leq a \leq r-1$;*
- *soit isomorphes à $\mathcal{R}_{\frac{r+1}{2}}$ (si r est impair).*

De plus, si H_0 n'est pas une étoile, on rencontre au moins un des deux derniers cas.

De ce résultat, nous déduisons la borne suivante :

Théorème 6.10 (Auger, Charon, Hudry, Lobstein [12] & appendice J). *Un graphe connexe d'ordre n qui n'est pas un graphe complet admet au plus*

$$r(n) = \frac{1}{2} + \sqrt{\frac{1}{4} + 2n}$$

racines carrées dans $\mathcal{G}_6$. Si $n \geq 11$, alors G admet exactement $r(n)$ racines carrées dans $\mathcal{G}_6$ si et seulement $r(n)$ est entier et G est isomorphe au carré de $\mathcal{R}_{2,r(n)-2}$.

Par exemple, le graphe $(\mathcal{R}_{2,7})^2$ est celui qui possède le plus de racines carrées dans $\mathcal{G}_6$ parmi tous les graphes non complets à 36 sommets, à savoir 9.

Troisième partie

Annexes

Annexe A

Edge number, Minimum Degree, Maximum Independent Set, Radius and Diameter in Twin-Free Graphs

David Auger [1], **Irène Charon** [1],
Iiro Honkala [2], **Olivier Hudry** [1], **Antoine Lobstein** [3]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr, honkala@utu.fi*

Abstract

Consider a connected, undirected graph $G = (V, E)$ and an integer $r \geq 1$; for any vertex $v \in V$, let $B_r(v)$ denote the ball of radius r centred at v, i.e., the set of all vertices linked to v by a path consisting of at most r edges. If for all vertices $v \in V$, the sets $B_r(v)$ are different, then we say that G is r-twin-free.

In r-twin-free graphs, we prolong the study of the extremal values that can be achieved by the main classical parameters in graph theory, and investigate here the number of edges, the minimum degree, the size of a maximum independent set, as well as radius and diameter.

Keywords : Graph theory, identifying code, twins, twin-free graph, identifiable graph, minimum degree, maximum independent set, maximum stable set, radius, diameter.
2000 *Mathematics Subject Classification :* 05C99, 05C35.

1. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Department of Mathematics, University of Turku, 20014 Turku, Finland. The third author's research is supported by the Academy of Finland under grant 210280.

3. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

A.1 Introduction

A.1.1 Definitions and notation

Given an integer $r \geq 1$ and a graph $G = (V, E)$ which is connected, undirected and finite, we define $B_r(v)$, the *ball* of radius r centred at $v \in V$, by

$$B_r(v) = \{x \in V : d(x, v) \leq r\},$$

where $d(x, v)$ denotes the number of edges in any shortest path between v and x.

Whenever $d(x, v) \leq r$, we say that x and v *r-cover* each other (or simply *cover* if there is no ambiguity). A set $X \subseteq V$ covers a set $Y \subseteq V$ if every vertex in Y is covered by at least one vertex in X.

Two distinct vertices $v_1, v_2 \in V$ such that $B_r(v_1) = B_r(v_2)$ are called *r-twins* or *twins.* If G has no r-twins, that is, if

$$\forall\, v_1, v_2 \in V \text{ with } v_1 \neq v_2,\ B_r(v_1) \neq B_r(v_2), \tag{A.1}$$

then G is said to be *r-twin-free* or *twin-free.*

Twin-free graphs are of interest and have been studied because they are strongly connected with *identifying codes* [65], which we now define.

A *code* C is a nonempty set of vertices, and its elements are called *codewords.* For each vertex $v \in V$, we denote by

$$K_{C,r}(v) = C \cap B_r(v)$$

the set of codewords which r-cover v. Two vertices v_1 and v_2 with $K_{C,r}(v_1) \neq K_{C,r}(v_2)$ are said to be *r-separated*, or *separated*, by code C.

A code C is called *r-identifying*, or *identifying*, if the sets $K_{C,r}(v), v \in V$, are all nonempty and distinct [65]. In other words, all vertices must be covered and pairwise separated by C.

Remark A.1. *The graph $G = (V, E)$ admits at least one r-identifying code if, and only if, it is r-twin-free.*

Indeed, if for all $v_1, v_2 \in V$, $B_r(v_1)$ and $B_r(v_2)$ are different, then $C = V$ is r-identifying. Conversely, if for some $v_1, v_2 \in V$, $B_r(v_1) = B_r(v_2)$, then for any code $C \subseteq V$, we have $K_{C,r}(v_1) = K_{C,r}(v_2)$. This is why r-twin-free graphs are also called r-identifiable.

For instance, there is no r-identifying code in a complete graph, or clique*, with at least two vertices.*

A connected r-twin-free graph has one vertex or at least $2r + 1$ vertices (cf. Proposition A.2).

In the following, n will denote the number of vertices in G. For any integer $q > 0$, P_q will denote the path on q vertices, and the length of P_q will be equal to $q - 1$, its number of edges. Moreover, if v_1, v_2, …, v_q denote the vertices in P_q, we shall assume that these vertices are numbered in such a way that the edges in P_q are $\{v_i, v_{i+1}\}$ for $1 \leq i < q$. The cycle of length q (with q vertices and q edges, $q \geq 3$), consisting of P_q to which we add the edge $\{v_q, v_1\}$, will be denoted by $\mathcal{C}_q$.

The *radius* of a graph $G = (V, E)$ is defined by $\rho(G) = \min_{x \in V} \max_{y \in V} d(x, y)$. The *diameter*, $\delta(G)$, of a graph G is the largest possible distance between two vertices. A subset of vertices is said to be *independent*, or *stable*, if no two of its vertices are adjacent in G.

The following graph will be used in the sequel : we shall call it the *star*, and it consists of n vertices $0, 1, \ldots, n-1$, and $n-1$ edges $\{0, i\}$, $1 \leq i \leq n-1$. The graph with n vertices and all possible $n(n-1)/2$ edges is called the *clique* and denoted by K_n.

Given a connected graph $G = (V, E)$ and an integer $\ell \geq 2$, the *ℓ-transitive closure* of G, denoted by $G^{[\ell]}$, is the graph with vertex set $V^{[\ell]} = V$ and edge set $E^{[\ell]} = \{\{i, j\} : i, j \in V, 0 < d(i, j) \leq \ell\}$, where $d(.,.)$ refers to the distance in G. Given two graphs $G = (V, E)$ and $G^* = (V^*, E^*)$, the *product* $G \,\square\, G^*$ of G and G^* has vertex set $V \times V^*$ and edge set

$$\{\{(u, u^*), (v, v^*)\} : (u = v \text{ and } \{u^*, v^*\} \in V^*) \text{ or } (\{u, v\} \in V \text{ and } u^* = v^*)\}.$$

Finally, we recall that a *matching* in $G = (V, E)$ is any subset E_0 of E such that no two edges in E_0 have a vertex in common.

A.1.2 Illustration

The motivations for identifying codes come, for instance, from fault diagnosis in a multiprocessor system [65]. Such a system can be modeled as a graph $G = (V, E)$ where the vertices are the processors and the edges are the links between the processors. Assuming that at most one of the processors is malfunctioning, we wish to test the system and locate the faulty processor. For this purpose, some processors (constituting the code C) will be selected and assigned the task of testing their r-neighbourhoods (i.e., the vertices at distance at most r) : to every codeword $c \in C$ we ask the query "Is there a faulty vertex in $B_r(c)$?". If all the answers are NO, then, because C r-covers V, we know that there is no malfunctioning processor ; in case of YES answers, we know that there is one malfunctioning processor and we can locate it, because there is a unique vertex $v \in V$ such that $\{c \in C : c \text{ answered YES}\} = K_{C,r}(v)$.

Identifying codes were introduced in [65], and since then they have grown into a combinatorial and graph-theoretical topic of their own, studied in a large number of various papers, investigating particular graphs or families of graphs (such as certain infinite regular grids, trees, chains, cycles, planar graphs, or the hypercube), dealing with complexity issues, using heuristics such as the noising methods for the construction of good codes, or defining new notions such as robustness or $(r, \leq \ell)$-identifying codes, ... For a bibliography on identifying codes and related concepts, with already more than 150 items, see [73].

Therefore, it is quite natural to study some of the parameters of twin-free graphs, since these graphs, and only these graphs, admit identifying codes.

A.1.3 Scope of the paper

We intend to investigate the extremal values that some parameters, classical in graph theory, can achieve in any connected twin-free graph G. More precisely, for a parameter p (such as : the number of edges, the maximum degree, the diameter, ...), we fix r and search for the smallest value, $f_r(p)$, that this parameter can reach in G, or we fix r and n and search for the smallest and largest values, $f_{r,n}(p)$ and $F_{r,n}(p)$ respectively, that this parameter can reach in G, n being the number of vertices in G :

$$f_r(p) = \min\{p(G) : G \in \mathcal{G}_r\},$$

where $\mathcal{G}_r = \{G : G \text{ connected, } r\text{-twin-free with at least } 2r + 1 \text{ vertices}\}$;

$$f_{r,n}(p) = \min\{p(G) : G \in \mathcal{G}_{r,n}\} \quad \text{and} \quad F_{r,n}(p) = \max\{p(G) : G \in \mathcal{G}_{r,n}\},$$

where $\mathcal{G}_{r,n} = \{G : G \text{ connected, } r\text{-twin-free with } n \geq 2r + 1 \text{ vertices}\}$.

The function $F_r(p) = \max\{p(G) : G \in \mathcal{G}_r\}$ would present much less interest, since, for the parameters that we deal with, F_r is not bounded by above.

In this paper, we are interested in the following five parameters :

– number of edges, ε,
– minimum degree, $\Delta_{\min}$,
– size α of a maximum stable set,
– radius, ρ,
– diameter, δ,

and study, for each of them, the functions f_r, $f_{r,n}$, and $F_{r,n}$. Each of the Sections A.2–A.6 deals with one parameter ; at the beginning of each section, for comparison, the extremal values of the parameter are given for connected graphs. Section A.7 recapitulates our results in three tables ($r = 1$, $r = 2$, $r \geq 3$). Note that we have all the exact values when $r = 1$.

The same study was run in [30] for the following four parameters : number of vertices, minimum size of an r-identifying code, r-domination number, maximum size of a clique, ω, and in [31] for the maximum degree, $\Delta_{\max}$.

In the case $r = 1$, the number of edges had already been investigated in [74, Sec. 4.1.2], and the minimum degree in [70, 53].

Of course, some of these parameters are connected in some way. For instance, a graph containing a large clique will have a quite large number of edges (cf. Lemma A.7). More interestingly, in Theorems A.22 and A.25, we use the very same construction, which gives a graph with a rather large minimum degree and a rather small maximum stable set, and this construction also has a rather small diameter, see Remark A.35 ; in Theorem A.34, we use two constructions in order to obtain the smallest possible diameter, and these constructions also contain a rather small maximum stable set, see Remark A.26. And in Lemma A.20, we obtain an inequality linking n, $f_{\lfloor \frac{r}{2} \rfloor,n}(\alpha)$ and $F_{r,n}(\Delta_{\min})$.

A.2 The number of edges

Of course, in any connected graph with n vertices, the number of edges lies between $n - 1$ (trees) and $n(n - 1)/2$ (clique).

The following proposition, together with the connectivity of the graphs considered, will immediately give Proposition A.3.

Proposition A.2. *[36, 74, 29] Let $r \geq 1$ and G be any connected r-twin-free graph with at least two vertices. Then we have : $n \geq 2r + 1$. Moreover, P_{2r+1} is the only connected r-twin-free graph with $2r + 1$ vertices.*

Proposition A.3. *For all $r \geq 1$, we have : $f_r(\varepsilon) = 2r$.*

Proposition A.4. *For all $r \geq 1$ and $n \geq 2r + 1$, we have : $f_{r,n}(\varepsilon) = n - 1$.*

Démonstration. Let $r \geq 1$ and $n \geq 2r+1$; because we consider connected graphs, $f_{r,n}(\varepsilon) \geq n - 1$. On the other hand, the fact that any path P_n, $n \geq 2r + 1$, is r-twin-free shows that $f_{r,n}(\varepsilon) = n - 1$. □

More interesting and difficult is the study of $F_{r,n}(\varepsilon)$, although the exact value is already known for $r = 1$.

Proposition A.5. *[74] Let G be any connected 1-twin-free graph with ε edges and n vertices, $n \geq 3$. Then we have : $\varepsilon \leq \frac{n(n-1)}{2} - \lfloor \frac{n}{2} \rfloor$. Moreover, there is only one connected 1-twin-free graph with this number of edges : the clique minus a maximum matching.*

Corollary A.6. *For all $n \geq 3$, we have : $F_{1,n}(\varepsilon) = \frac{n(n-1)}{2} - \lfloor \frac{n}{2} \rfloor$.*

The following result is easy but will be helpful.

Lemma A.7. *For all* $r \geq 1$ *and* $n \geq 2r+1$*, we have :*

$$\frac{F_{r,n}(\omega) \times (F_{r,n}(\omega) - 1)}{2} + n - F_{r,n}(\omega) \leq F_{r,n}(\varepsilon) \leq \frac{n \times F_{r,n}(\Delta_{\max})}{2}, \quad \text{(A.2)}$$

where ω *stands for the maximum size of a clique and* $\Delta_{\max}$ *for the maximum degree.*

Démonstration. The upper bound is immediate. For the lower bound, construct an r-twin-free graph with n vertices, containing a clique of size $F_{r,n}(\omega)$; in addition to the edges inside the clique, there must be at least $n - F_{r,n}(\omega)$ edges in order to connect the remaining vertices. □

The lower bound is rough : we only used the fact that the graph is connected, not that it is r-twin-free. However, we believe that this could bring only a small improvement, through heavy complications.

The previous lemma, together with the following two theorems from [31] and the forthcoming [30], gives the lower and upper bounds of Corollary A.10 for $F_{r,n}(\varepsilon)$.

Theorem A.8. *[30, Th. 27] For all* $r \geq 2$ *and* $n \geq 2r+1$*, we have :*

$$k_\omega \leq F_{r,n}(\omega) \leq k_\omega + 1,$$

where k_ω *is the largest integer such that* $k_\omega + r\lceil \log_2 k_\omega \rceil \leq n-1$*.*

Actually, [30, Th. 27] states when $F_{r,n}(\omega) = k_\omega$ and when $F_{r,n}(\omega) = k_\omega + 1$, but this is not very relevant here.

Theorem A.9. *[31, Th. 3] For all* $r \geq 2$ *and* $n \geq 2r+1$*, we have :*

$$F_{r,n}(\Delta_{\max}) \leq k_\Delta, \quad \text{(A.3)}$$

where k_Δ *is the largest integer such that*

$$k_\Delta + (r-2)\lceil \log_3(k_\Delta + 1) \rceil + \lceil \log_2(k_\Delta + 1) \rceil \leq n-1.$$

Corollary A.10. *For all* $r \geq 2$ *and* $n \geq 2r+1$*, we have :*

$$\frac{k_\omega \times (k_\omega - 1)}{2} + n - k_\omega \leq F_{r,n}(\varepsilon) \leq \frac{n \times k_\Delta}{2}, \quad \text{(A.4)}$$

where k_ω *and* k_Δ *are defined in Theorems A.8 and A.9, respectively.*

For $r > 2$ and with conditions on n and r, the upper bound in (A.4) can be improved using the following lemma and corollary ; from now on, we set

$$\ell = \log_2 3 \quad (\ell \approx 1.585).$$

Lemma A.11. *[31, proof of Th. 3] For* $r \geq 2$ *and* $n \geq 2r+1$*, let* $G = (V, E)$ *be any connected* r*-twin-free graph with* n *vertices, and let* v *be any vertex in* V*, the degree of which we denote by* $\deg(v)$*. Let also* V_0 *be the set of vertices adjacent to* v*, and, for* $i = 1, 2, \ldots, r, \ldots$*,* $V_i = \{x \in V \setminus \{v\} : d(x, V_0) = i\}$*, where as usual* $d(x, V_0)$ *is the smallest distance between* x *and the vertices in* V_0*. Then for* $i = 1, \ldots, r-1$*,*

$$|V_i| \geq \log_3(\deg(v) + 1) \text{ and } |V_{r-1}| + |V_r| \geq \log_2(\deg(v) + 1).$$

Moreover, the vertices V_i*,* $i \geq 0$*, partition* $V \setminus \{v\}$ *and there is no edge between two non-consecutive sets* V_i*.*

Corollary A.12. *For $r > 2$ and $n \geq 2r+1$, let $G = (V, E)$ be any r-twin-free graph with n vertices, and let v be a vertex whose degree* $\deg(v)$ *is maximum in G. Then the number of edges in G is at most*

$$\begin{cases} (a)\ \frac{n\deg(v)}{2} & \text{if } \deg(v) \leq \frac{n-2}{2}, \\ (b)\ \frac{n\deg(v)-(2\deg(v)-n+2)(r-3+\ell)\log_3\deg(v)}{2} & \text{if } \deg(v) \geq \frac{n-2}{2}. \end{cases} \tag{A.5}$$

Démonstration. In all cases, since v has maximum degree, we have $|E| \leq n\deg(v)/2$. Using the notation of Lemma A.11, and letting $t = \sum_{i\geq 2}|V_i|$, we know that $t \geq (r-3)\log_3(\deg(v)+1) + \log_2(\deg(v)+1) = (r-3+\ell)\log_3(\deg(v)+1)$, because $r > 2$. For simplicity, we slightly weaken this result, to obtain

$$t \geq (r-3+\ell)\log_3\deg(v). \tag{A.6}$$

Because there is no edge between two non-consecutive sets V_i, a vertex in V_i, $i \geq 2$, cannot be adjacent to v nor to any vertex in V_0 — and is not adjacent to itself ; so its degree is at most $n - |V_0| - 2 = n - \deg(v) - 2$. The fact that v has been chosen with maximum degree leads to :

$$\begin{aligned} 2|E| &= \sum_{w\in V}\deg(w) && \text{(A.7)} \\ &\leq t(n-\deg(v)-2) + (n-t)\deg(v) && \text{(A.8)} \\ &= n\deg(v) - t(2\deg(v)-n+2). && \text{(A.9)} \end{aligned}$$

If $\deg(v) \geq \frac{n-2}{2}$, then $2\deg(v) - n + 2 \geq 0$ and we can combine (A.6) and (A.9), which leads to

$$2|E| \leq n\deg(v) - (r-3+\ell)(2\deg(v)-n+2)\log_3\deg(v). \tag{A.10}$$

□

We shall need to know the behaviour of the right-hand side of inequality (A.10), and to do so, we need a condition on r ; note that we did not try to get the best possible condition.

Lemma A.13. *Let $g_{r,n}(x) = nx - (r-3+\ell)(2x-n+2)\log_3 x$. If $2 < r \leq n/3\log_3 n$, then $g_{r,n}(x)$ in an increasing function for x between 1 and $n-1$.*

Démonstration. Computations show that $g'_{r,n}(x)$ can read :

$$g'_{r,n}(x) = (n - 3(r-3+\ell)\log_3 x) - (r-3+\ell)\frac{2x-n+2-x\ln x}{x\ln 3}.$$

Now $x(2-\ln x) - n + 2 < 0$ when $1 \leq x < n-1$, and $r - 3 + \ell > 0$, so

$$g'_{r,n}(x) > n - 3(r-3+\ell)\log_3 x > n - 3r\log_3 x > n - 3r\log_3 n \geq 0.$$

□

Till the end of this section, we assume, for simplicity, that r is "small" with respect to n, so that we can use the following approximations for k_ω and k_Δ :

$$k_\omega \approx n - r\log_2 n \quad [30, \text{Prop. } 28], \tag{A.11}$$

$$k_\Delta \approx n - (r-2+\ell)\log_3 n \quad [31, (13)]. \tag{A.12}$$

Moreover, we can also use $F_{r,n}(\Delta_{\max}) \gtrsim n - r\log_3 n$ [31, (12)], which shows that in this case, k_Δ is a not too bad estimate of $F_{r,n}(\Delta_{\max})$.

Corollary A.14. *Assume that $r > 2$ is small compared to n. Then*

$$F_{r,n}(\varepsilon) \leq \frac{nk_\Delta - (2k_\Delta - n + 2)(r - 3 + \ell)\log_3 k_\Delta}{2} \tag{A.13}$$

$$\approx \frac{n^2}{2} - n\,\ell^{-1}(r + \ell - 5/2)\log_2 n. \tag{A.14}$$

Démonstration. Let $G = (V, E)$ be any connected r-twin-free graph with n vertices and maximum degree $\Delta(G)$.

If $\Delta(G) \leq \frac{n-2}{2}$, then by Corollary A.12 we know that G has at most $\frac{n(n-2)}{4}$ edges.

If $\Delta(G) \geq \frac{n-2}{2}$, then by the same corollary, G has at most $\frac{1}{2}g_{r,n}(\Delta(G))$ edges, with $g_{r,n}$ defined in Lemma A.13. Now, using (A.3), (A.12) and the fact that $g_{r,n}$ is increasing, we have

$$\begin{aligned}\tfrac{1}{2}g_{r,n}(\Delta(G)) &\leq \tfrac{1}{2}g_{r,n}(F_{r,n}(\Delta_{\max}))\\ &\leq \tfrac{1}{2}g_{r,n}(k_\Delta) = \tfrac{nk_\Delta-(2k_\Delta-n+2)(r-3+\ell)\log_3 k_\Delta}{2}\\ &\approx \tfrac{1}{2}g_{r,n}(n - (r-2+\ell)\log_3 n),\end{aligned}$$

which gives approximately, for r sufficiently small compared to n,

$$\frac{n^2}{2} - n\,\ell^{-1}(r + \ell - 5/2)\log_2 n,$$

a quantity which is greater than $\frac{n(n-2)}{4}$ for small r.

So among all connected r-twin-free graphs with n vertices, the number of edges is at most

$$\frac{nk_\Delta - (2k_\Delta - n + 2)(r - 3 + \ell)\log_3 k_\Delta}{2} \approx \frac{n^2}{2} - n\,\ell^{-1}(r + \ell - 5/2)\log_2 n.$$

□

Now, we estimate the gap between the lower bound given by (A.4) and the upper bound given by (A.2) (case $r = 2$) or by (A.14) (case $r > 2$, r "small").

First, we consider the case $r = 2$. Theorem 5 in [31] states that

$$F_{2,n}(\Delta_{\max}) = n - p_1 - 2,$$

where p_1 is such that $2^{p_1} + p_1 + 1 \leq n \leq 2^{p_1+1} + p_1 + 1$, so that p_1 is close to $\log_2 n$. On the other hand, surveying the three possible cases for n :

$$n = 2^{p_2} + 2p_2 + 1,\; n = 2^{p_2} + 2p_2 + 2, \;\text{ and }\; 2^{p_2} + 2p_2 + 3 \leq n \leq 2^{p_2+1} + 2p_2 + 2,$$

we can see from the definition of k_ω in Theorem A.8 that $k_\omega = n-2p_2-1$, $k_\omega = n-2p_2-2$, and $k_\omega = n - 2p_2 - 3$, respectively, with p_2 close to $\log_2 n$.

We can therefore approximate the lower bound in (A.4) by replacing k_ω by $n - 2p_2$, and the upper bound in (A.2) by replacing $F_{2,n}(\Delta_{\max})$ by $n - p_1$, obtaining

$$\frac{(n-2p_2)^2}{2} + 3p_2 - \frac{n}{2} \lesssim F_{2,n}(\varepsilon) \lesssim \frac{n(n-p_1)}{2}\ ,$$

which, when n grows, yields

$$\frac{n^2}{2} - 2n\log_2 n \lesssim F_{2,n}(\varepsilon) \lesssim \frac{n^2}{2} - \frac{n\log_2 n}{2}, \tag{A.15}$$

the gap between lower and upper bounds being $\frac{3}{2}n\log_2 n$.

Now we assume that $r > 2$ is fixed or grows slowly with respect to n ; in this case, plugging (A.11) into (A.4) and using (A.14) ultimately lead to

$$\frac{n^2}{2} - rn\log_2 n \lesssim F_{r,n}(\varepsilon) \lesssim \frac{n^2}{2} - n\,\ell^{-1}(r+\ell-5/2)\log_2 n,$$

which, with the numerical value of ℓ, reads :

$$n^2/2 - rn\log_2 n \lesssim F_{r,n}(\varepsilon) \lesssim n^2/2 - 0.63\,(r-0.915)\,n\log_2 n. \tag{A.16}$$

Here, the gap between the bounds is approximately $(0.37r + 0.58)n\log_2 n$.

To have an intuitive grasp of the lower bound in (A.16), the reader must imagine a construction very similar to the one in Figure A.II, with a clique in place of the stable set V_1.

Example A.15. *Let* $r = 10$, $n = 59\,153$ *; then* $k_\omega = 58\,992$ *and* $k_\Delta = 59\,048$*. By [30, Th. 27], we obtain* $F_{10,59\,153}(\omega) = 58\,993$*. The fact that* $\frac{59\,153\times 59\,151}{4} \leq \frac{1}{2}g_{10,59\,153}(59\,048)$ *shows, see proof of Corollary A.14, that (A.13) is a valid upper bound. So, together with the lower bound in (A.4), we obtain*

$$1\,740\,057\,688 \leq F_{10,59\,153}(\varepsilon) \leq 1\,743\,902\,958,$$

with a difference of 0.2% *between lower and upper bounds. The estimates given by (A.16) lead to* $1\,740\,161\,000 \leq F_{10,59\,153}(\varepsilon) \leq 1\,744\,172\,000$.

A.3 The minimum degree, $\Delta_{\min}$

In any connected graph with n vertices, the minimum degree is obviously comprised between 1 (e.g., trees) and $n-1$ (clique).

The study of $f_r(\Delta_{\min})$ and $f_{r,n}(\Delta_{\min})$ is easy. Note that the case $r = 1$ is mentioned in [70], as part of a more general result.

Theorem A.16. *For all* $r \geq 1$*, we have :* $f_r(\Delta_{\min}) = 1$*. For all* $r \geq 1$ *and* $n \geq 2r+1$*, we have :* $f_{r,n}(\Delta_{\min}) = 1$.

Démonstration. The minimum degree is at least one because we consider only connected graphs. It is equal to one in every r-twin-free path. □

More challenging is the function $F_{r,n}(\Delta_{\min})$. We have the exact values for $r = 1$, $r = 2$.

Theorem A.17. *For all* $n \geq 3$*, we have :* $F_{1,n}(\Delta_{\min}) = n-2$.

Démonstration. Since the clique with $n \geq 2$ vertices is not twin-free, we see that $n-2$ is an upper bound. The clique with n vertices minus a maximum matching is a connected 1-twin-free graph (see Proposition A.5) with minimum degree equal to $n-2$. □

Theorem A.18. *For all* $n \geq 5$*, we have :* $F_{2,n}(\Delta_{\min}) = \lfloor\frac{n-2}{2}\rfloor$.

Démonstration. In order to prove the lower bound, we construct a complete bipartite graph deprived of a maximum matching : let $n = 2p$ or $2p+1$, and $G = (V, E)$, with $V = V_1 \cup V_2$, $V_1 = \{v_1, \ldots, v_p\}$, $V_2 = \{v_{p+1}, \ldots, v_n\}$, $E = \{\{v_i, v_j\} : 1 \leq i \leq p,\ p+1 \leq j \leq n\} \setminus E_0$, where E_0 is any matching with p edges. It is not difficult to see that G is connected and 2-twin-free, and has minimum degree $p-1$, so that $F_{2,n}(\Delta_{\min}) \geq p - 1 = \lfloor \frac{n-2}{2} \rfloor$.

On the other hand, take any connected 2-twin-free graph, $G = (V, E)$, with n vertices, $n \geq 5$. Consider two vertices $x, y \in V$ at maximum distance from one another in G. Necessarily, $d(x, y) \geq 3$ (otherwise, condition (A.1) with $r = 2$ could not be fulfilled — see also the proof of Theorem A.33 on the diameter). Let $\Delta(G)$ be the minimum degree in G. There are at least $\Delta(G)$ vertices adjacent to x, at least $\Delta(G)$ vertices adjacent to y, and a vertex adjacent to x cannot be adjacent to y, and vice versa. So, together with x and y, there are at least $2 + 2\Delta(G)$ vertices in G, and $\Delta(G) \leq \frac{n-2}{2}$. □

We now give two lemmas, the latter establishing an interesting relation between maximum stable set and minimum degree, and a strong result from [45]. They will be used in the proof of Theorem A.22 on $F_{r,n}(\Delta_{\min})$ for $r > 2$.

Lemma A.19. *Let $G = (V, E)$ be a connected graph with n vertices. For all $r, s, t \geq 1$ and $n \geq 2r + 1$, if $r = st$ and G is r-twin-free, then the s-transitive closure of G, $G^{[s]}$, is t-twin-free.*

Démonstration. In this proof as well as in the proof of Lemma A.20, to avoid ambiguities we shall use the notation $d_G(x, y)$ (respectively, $d_{G^{[s]}}(x, y)$) when the vertices x, y are considered in G (respectively, in $G^{[s]}$).

By definition, $0 < d_{G^{[s]}}(x, y) \leq 1$ if and only if $0 < d_G(x, y) \leq s$, which in turn implies that $0 < d_{G^{[s]}}(x, y) \leq t$ if and only if $0 < d_G(x, y) \leq st = r$. □

Lemma A.20. *For all $r \geq 2$ and $n \geq 2r + 1$, we have :*

$$f_{\lfloor \frac{r}{2} \rfloor, n}(\alpha) \times (1 + F_{r,n}(\Delta_{\min})) \leq n. \tag{A.17}$$

Démonstration. Let $G = (V, E)$ be any connected graph with n vertices, minimum degree $\Delta(G)$ and no r-twins. First we assume that r is even. Then, by Lemma A.19, $G^{[2]}$ is $\frac{r}{2}$-twin-free. Let $\mathcal{S} = \{s_1, \ldots, s_m\}$ be a maximum stable set in $G^{[2]}$:

$$m \geq f_{\frac{r}{2}, n}(\alpha).$$

We shall see in Theorems A.24 and A.25 that we always have $f_{r,n}(\alpha) \geq 2$, but anyway the end of the following argument also works if $m = 1$.

For $i \neq j$, let s_i, s_j be two elements of $\mathcal{S}$, and let $B_{G,1}(x) = \{y \in V : d_G(x, y) \leq 1\}$ for $x \in V$. We have the following implications :

$$\begin{aligned} s_i, s_j \in \mathcal{S} \quad &\Longrightarrow d_{G^{[2]}}(s_i, s_j) \geq 2 \Longrightarrow d_G(s_i, s_j) \geq 3 \\ &\Longrightarrow B_{G,1}(s_i) \cap B_{G,1}(s_j) = \emptyset \Longrightarrow \textstyle\sum_{p=1}^{m} |B_{G,1}(s_p)| \leq n \\ &\Longrightarrow m \times (1 + \Delta(G)) \leq n \Longrightarrow f_{\frac{r}{2}, n}(\alpha) \times (1 + \Delta(G)) \leq n. \end{aligned}$$

If r is odd ($r \geq 3$), then, since G is *a fortiori* $(r-1)$-twin-free, we can repeat the above argument with $r - 1$ and obtain $f_{\frac{r-1}{2}, n}(\alpha) \times (1 + \Delta(G)) \leq n$.

Therefore, $f_{\lfloor \frac{r}{2} \rfloor, n}(\alpha) \times (1 + \Delta(G)) \leq n$. Since this is true for all r-twin-free graphs G with n vertices and minimum degree $\Delta(G)$, inequality (A.17) is verified. □

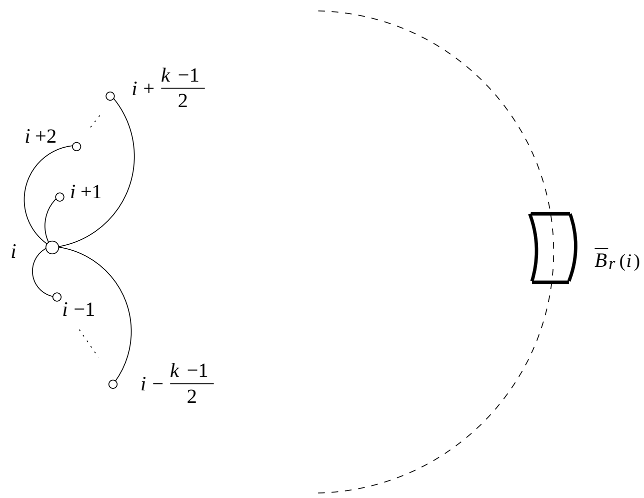


Figure A.I – *A partial representation of the graph G constructed in the proof of Theorem A.22, for odd k : the neighbours of i and the vertices not r-covering i.*

Theorem A.21. *[45] Let* $G = (V, E)$ *be a graph with* n *vertices and minimum degree* $\Delta(G) \geq 2$*. Then*

$$\rho(G) = \min_{x \in V} \max_{y \in V} d(x, y) \leq \frac{3n-3}{2\Delta(G)+1} + 5.$$

Theorem A.22. *For all* $r \geq 3$*, we have :* $F_{r,2r+1}(\Delta_{\min}) = 1$*.*
For all $r \geq 3$ *and* $n \geq 2r+2$*, let* $k = \lfloor \frac{n-2}{r} \rfloor$*. Then we have :*

$$k - 1 \leq F_{r,n}(\Delta_{\min}) \text{ if } k \text{ is odd, and } k \leq F_{r,n}(\Delta_{\min}) \text{ if } k \text{ is even.} \tag{A.18}$$

For $r = 3, 4, 5$ *and* $n \geq 2r+2$*,* $F_{r,n}(\Delta_{\min}) \leq \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1$ *; for all* $r \geq 6$ *and* $n \geq 2r+2$*,*

$$F_{r,n}(\Delta_{\min}) \leq \min\left\{ \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1, \; \frac{3n - r + 2}{2(r-5)} \right\}.$$

Démonstration. If $n = 2r+1$, then necessarily $G = P_{2r+1}$ (cf. Proposition A.2) and G has minimum degree one, so we assume from now on that $n \geq 2r+2$. An equivalent definition of k reads $n-2 = kr+\beta$, with $0 \leq \beta < r$, $k \geq 2$. The lower bounds involving k are obtained through the following construction : the graph G has vertex set $V = \{0, 1, \dots, n-1\}$ and edge set $E = \{\{i, i+j \mod n\}\}$, where $i \in V$,

$$j \in \begin{cases} \{-k/2, \dots, -1, 1, \dots, k/2\} & \text{if } k \text{ is even,} \\ \{-(k-1)/2, \dots, -1, 1, \dots, (k-1)/2\} & \text{if } k \text{ is odd,} \end{cases}$$

cf. Figure A.I. In both cases, the graph is regular, with degree either k (if k is even) or $k-1$ (if k is odd).

We now prove that G is r-twin-free, which will establish (A.18). For further use of this construction, note that the proof also works for $r = 1$ and $r = 2$. If k is even, then for every $i \in V$,

$$B_r(i) = \{i - kr/2, \dots, i-1, i, i+1, \dots, i + kr/2\},$$

where operations are carried modulo n. Now let $\overline{B}_r(i) = V \setminus B_r(i)$: if one pictures G as the $\frac{k}{2}$- or $\frac{k-1}{2}$-transitive closure of the cycle $\mathcal{C}_n$, then $\overline{B}_r(i)$, which is the set of vertices not

r-covering i, is an interval situated opposite i, and it is quite straightforward to see that this set contains $n-1-kr=\beta+1$ elements, a number which ranges between 1 and r. If k is odd, then, similarly, $\overline{B}_r(i)$ contains $n-1-kr+r=\beta+1+r$ elements, a number lying between $r+1$ and $2r$. As a consequence, in both cases, all the sets $B_r(i)$ are different.

The upper bound $\frac{n}{\lfloor\frac{r}{2}\rfloor+1}-1$ comes directly by combining inequality (A.17) with the left-hand side of the forthcoming inequality (A.19), applied to $\lfloor\frac{r}{2}\rfloor$, with $r\geq 4$; if $r=3$, we combine (A.17) and $f_{1,n}(\alpha)=2$ from Theorem A.24.

Now let $G=(V,E)$ be any connected r-twin-free graph with n vertices, minimum degree $\Delta(G)\geq 2$ and radius $\rho(G)$. Then $r\leq\rho(G)$ (see the proof of Theorem A.30), and, using Theorem A.21, we obtain $r\leq\frac{3n-3}{2\Delta(G)+1}+5$, which leads to the upper bound $\frac{3n-r+2}{2(r-5)}$, when $r\geq 6$. □

Let $\mathcal{A}=\frac{n}{\lfloor\frac{r}{2}\rfloor+1}-1$ and $\mathcal{B}=\frac{3n-r+2}{2(r-5)}$; when $r\geq 23$ is odd or $r\geq 26$ is even, the coefficient $\frac{3}{2(r-5)}$ of n in $\mathcal{B}$ is smaller than or equal to the coefficient $\frac{1}{\lfloor\frac{r}{2}\rfloor+1}$ of n in $\mathcal{A}$, and therefore $\mathcal{B}$ becomes better than $\mathcal{A}$.

Based on the values for $r=1$ and $r=2$, we venture to conjecture that the exact value is $\lfloor\frac{n-2}{r}\rfloor$ for all r, which would mean that the construction in Theorem A.22 is quasi-optimum.

A.4 The size of a maximum stable set, α

In any connected graph with n vertices, the maximum size of a stable set ranges between 1 (clique) and $n-1$ (star).

Thanks to a recent result, we know, for all $r\geq 1$, the exact value for $f_r(\alpha)$.

Theorem A.23. *For all $r\geq 1$, we have : $f_r(\alpha)=r+1$.*

Démonstration. Let G be any connected r-twin-free graph with n vertices. It was conjectured in [29], and it is now proved in [4],[5], that there exists a chordless path with $2r+1$ vertices in G. Taking in this path $r+1$ vertices constituting a stable set proves the lower bound.

The paths P_{2r+1}, P_{2r+2} and the cycles $\mathcal{C}_{2r+2}, \mathcal{C}_{2r+3}$ are connected r-twin-free graphs containing stable sets of maximum size $r+1$. □

We know $f_{r,n}(\alpha)$ when $r=1$, but we have only bounds in the general case.

Theorem A.24. *For all $n\geq 3$, we have : $f_{1,n}(\alpha)=2$.*

Démonstration. To prove that $f_{1,n}(\alpha)\geq 2$, use the lower bound of Theorem A.23 with $r=1$.

The upper bound comes from the clique with n vertices deprived of a maximum matching E_0 (which is 1-twin-free, see Proposition A.5) : any three vertices, which formed a triangle in the clique, cannot have been disconnected when removing the edges of E_0 ; therefore, they cannot constitute a stable set. □

Theorem A.25. *For all $r\geq 2$, we have : $f_{r,2r+1}(\alpha)=r+1$.*

For all $r\geq 2$ and $n\geq 2r+2$, let $k=\lfloor\frac{n-2}{r}\rfloor$. Then

$$r+1\leq f_{r,n}(\alpha)\leq\begin{cases}\text{(a) } \lfloor\frac{2n}{k+2}\rfloor & \text{if } k \text{ is even,}\\ \text{(b) } \lfloor\frac{2n}{k+1}\rfloor & \text{if } k \text{ is odd.}\end{cases}\tag{A.19}$$

In both cases, we have :

$$f_{r,n}(\alpha)\leq 2r.\tag{A.20}$$

Démonstration. See Theorem A.23 for the lower bound. The case $n = 2r+1$ is easy, so we turn to $n \geq 2r+2$. The two upper bounds are obtained through the graph G constructed in the proof of Theorem A.22. Since G is r-twin-free for all $r \geq 2$, all we have to establish is the size of a maximum stable set in G.

Consider a stable set $\mathcal{S}$ of G and an element s of $\mathcal{S}$. Following the very definition of G, no other vertex in $\mathcal{S}$ can belong to the set $\{s-k/2, \ldots, s-1, s, s+1, \ldots, s+k/2\}$ if k is even, and to the set $\{s-(k-1)/2, \ldots, s-1, s, s+1, \ldots, s+(k-1)/2\}$ if k is odd, with computations carried modulo n. Now a packing-type argument immediately gives that

$$|\mathcal{S}| \leq \left\lfloor \frac{n}{\frac{k}{2}+1} \right\rfloor \text{ if } k \text{ is even and } |\mathcal{S}| \leq \left\lfloor \frac{n}{\frac{k-1}{2}+1} \right\rfloor \text{ if } k \text{ is odd.}$$

From (A.19a) and (A.19b) we easily get the inequality

$$f_{r,n}(\alpha) \leq 2r\frac{n}{n-1},$$

and since n is at least $2r+2$, the right-hand side is always smaller than $2r+1$, which gives (A.20). □

Remark A.26. *It is not difficult to see that both constructions described in Cases 1 and 2 of the proof of Theorem A.34 also contain a maximum stable set of size at most $2r$. The construction used in Case 1 is liable to give an even better upper bound for $f_{r,n}(\alpha)$, but this would be valid only for $2r+2 \leq n \leq 4r+3$.*

Corollary A.27. *(a) We have : $f_{2,5}(\alpha) = 3$.*
(b) For all $n \geq 6$, let $k = \lfloor \frac{n-2}{2} \rfloor$. Then

$$f_{2,n}(\alpha) = 3 \textit{ if } k \textit{ is even, and } 3 \leq f_{2,n}(\alpha) \leq 4 \textit{ if } k \textit{ is odd.}$$

Démonstration. Simply apply the equality $n-2 = 2k+\beta$, $\beta \in \{0,1\}$, $k \geq 2$, to the two cases (A.19a) : k even, and (A.19b) : k odd. □

We now turn to the function $F_{r,n}(\alpha)$. When $r = 1$, the obvious upper bound $n-1$ is the exact value for $F_{1,n}(\alpha)$.

Theorem A.28. *For all $n \geq 3$, we have : $F_{1,n}(\alpha) = n-1$.*

Démonstration. Use the star, defined in the Introduction. □

Finally, we study the general case for $F_{r,n}(\alpha)$.

Theorem A.29. *For all $r \geq 2$ and $n \geq 2r+1$, we have :*

$$\max\left\{\left\lceil \frac{n}{2} \right\rceil, \; k + \lceil \log_2 k \rceil \left\lfloor \frac{r}{2} \right\rfloor\right\} \leq F_{r,n}(\alpha) \leq n-r,$$

where k is the largest integer such that $k + r\lceil \log_2 k \rceil \leq n-1$.

Démonstration. Let G be any connected r-twin-free graph with n vertices, and let $p = \lceil \log_2 k \rceil$. There exists in G a path with $2r+1$ vertices ([29] ; see also the proof of Theorem A.23, which contains a stronger statement), which shows that at least r vertices cannot belong to a stable set in G. The upper bound follows.

For the lower bound, either we use the path P_n, or we construct the following graph $G = (V, E)$ with $n \geq 2r+1$ vertices.

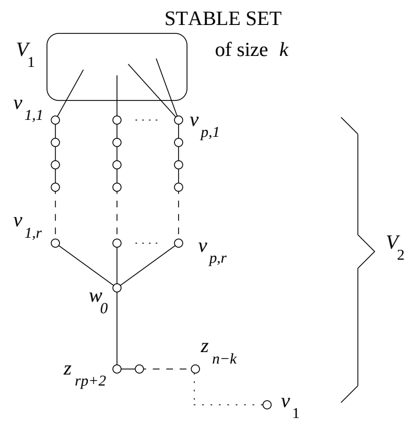


Figure A.II – *Partial representation of the graph constructed for Theorem A.29.*

If $2^{p-1} < k < 2^p$, then $p = \lceil \log_2 k \rceil = \lceil \log_2(k+1) \rceil$, $k + rp \le n-1$ and $(k+1)+rp > n-1$; therefore, we obtain :

$$rp + 1 = n - k. \tag{A.21}$$

If $k = 2^p$, then, since $(k+1) + r\lceil \log_2(k+1) \rceil = k+1+r(p+1) \ge n$, we obtain :

$$rp + 1 \le n - k \le rp + 1 + r. \tag{A.22}$$

Now let $G_1 = (V_1, E_1)$ be a stable set with k vertices : $|V_1| = k$ and $E_1 = \emptyset$. To V_1 we add a set V_2 of $n-k$ vertices, where

$$V_2 = \{w_0\} \cup \{v_{i,j} : 1 \le i \le p, 1 \le j \le r\} \cup \{z_j : rp+2 \le j \le n-k\},$$

see Figure A.II. Note that by (A.21) and (A.22), there are between 0 and r vertices z_j in V_2. We construct the edges $\{v_{i,j}, v_{i,j+1}\}$ for $1 \le i \le p$, $1 \le j \le r-1$, and $\{z_j, z_{j+1}\}$ for $rp+2 \le j \le n-k-1$, to which we add the edges $\{w_0, z_{rp+2}\}$ and $\{w_0, v_{i,r}\}$, $1 \le i \le p$.

Then we create new edges as follows : we link each of the k vertices in V_1 to a different subset of vertices taken among the vertices $v_{i,1}$, $1 \le i \le p$. This is possible because $k \le 2^p$, and this may include the empty set for a particular vertex $v_1 \in V_1$. Obviously, this can be done in such a way that G is connected, unless v_1 exists ; in this case, we may choose to link v_1 to z_{n-k}, or to w_0 if there is no vertex of type z_j, and we consider that $v_1 \in V_2 = V \setminus V_1$ and that $|V_1|$ becomes equal to $k-1$.

The graph G is r-twin-free : one can see that two vertices in V_1 are not r-twins, since by construction the vertices $v_{i,r}$, $1 \le i \le p$, all contribute differently to the balls of radius r centred at the vertices in the stable set ; one vertex in V_1 and one vertex in V_2 cannot be r-twins, because w_0 r-covers all the vertices in V_2, except maybe v_1, and no vertex in V_1, but obviously v_1 has no twins in V_1. Finally, we have to consider two vertices in V_2. If $p = 1$, then $k = 2$ and we have constructed the path P_n consisting of one vertex in V_1, r vertices $v_{1,i}$, the vertex w_0, at least $r-2$ vertices z_j, $r+2 \le j \le n-2$, and the vertex v_1, which we linked to z_{n-2} ; this path is r-twin-free since $n \ge 2r+1$.

If $p > 1$, we first show that w_0 and a vertex $v_{i,j}$ are not twins : this is so because w_0 is r-covered by $v_{h,r-j+1}$ $(h \ne i)$, $v_{i,j}$ is not ; next, $v_{i,j}$ and $v_{i',j'}$ are not twins : if $i = i'$, then $v_{h,r-j+1}$ $(h \ne i)$ r-covers $v_{i,j'}$, but not $v_{i,j}$; and if $i \ne i'$, then $v_{i',r-j+1}$ r-covers $v_{i',j'}$, but not $v_{i,j}$. Lastly it is easy to check that the vertices z_m and v_1 are twins neither between themselves, nor with w_0, nor with any vertex $v_{i,j}$, because the balls of radius r centred at the vertices z_m and v_1 either cut the branches $v_{i,1}, \ldots, v_{i,r}$ at different vertices $v_{i,j}$,

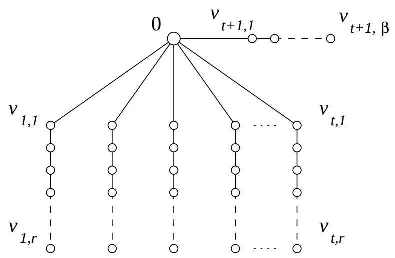


Figure A.III – *A generalization/subdivision of the star.*

$r \geq j > 1$, or do not cut any of these branches, whereas w_0 r-covers all the full branches and a vertex $v_{i,j}$ r-covers the whole branch i ; the vertices not cutting any branch are necessarily $z_{n-k} = z_{rp+1+r}$ or v_1, which are not twins either.

In this graph G, whether or not v_1 exists, there is a stable set of size $k + p\lfloor \frac{r}{2} \rfloor$. ☐

In [30], we observed that if r grows slowly with respect to n, e.g., if r is fixed and n grows, then k behaves approximately like $n - r \log_2 n$ and p like $\log_2 n$, and so $k + p\lfloor \frac{r}{2} \rfloor$ behaves like

$$n - \frac{r}{2} \log_2 n.$$

A.5 The radius, ρ

In any connected graph with n vertices, the radius ranges between 1 (clique, star) and $\lfloor \frac{n}{2} \rfloor$ (paths P_n, cycles $\mathcal{C}_n$).

Compared to other parameters, the study of the radius is easy, and we obtain the exact values for all three functions $f_r(\rho)$, $f_{r,n}(\rho)$, and $F_{r,n}(\rho)$.

Theorem A.30. *For all $r \geq 1$, we have : $f_r(\rho) = r$.*

Démonstration. The paths P_{2r+1} show that $f_r(\rho) \leq r$. For the lower bound, consider any r-twin-free graph $G = (V, E)$, and assume that there is a vertex $x \in V$ such that for all $y \in V$, we have $d(x, y) < r$. Then any neighbour z of x is within distance r from all vertices, and x and z are r-twins. This shows that the radius of G is at least r, and the inequality $f_r(\rho) \geq r$ follows. ☐

Theorem A.31. *For all $r \geq 1$ and $n \geq 2r + 1$, we have : $f_{r,n}(\rho) = r$.*

Démonstration. For the lower bound, see the previous theorem. For the upper bound, we use a graph $G = (V, E)$, called the *generalized star*, which is built as follows : first, we divide $n - 1$ by r, so $n = tr + \beta + 1$, with $t \geq 2$ and $0 \leq \beta < r$. Then we set $V = \{0\} \cup \{v_{i,j} : 1 \leq i \leq t, 1 \leq j \leq r\} \cup \{v_{t+1,j} : 1 \leq j \leq \beta\}$, and $E = \{\{0, v_{i,1}\}, \{v_{i,j}, v_{i,j+1}\} : 1 \leq i \leq t, 1 \leq j \leq r-1\} \cup \{\{0, v_{t+1,1}\}\} \cup \{\{v_{t+1,j}, v_{t+1,j+1}\} : 1 \leq j \leq \beta-1\}$; see Figure A.III. In other words, starting from 0, we have t "branches" with r edges, and one with β edges. This graph has n vertices, $n - 1$ edges, radius r and is r-twin-free. ☐

Theorem A.32. *For all $r \geq 1$ and $n \geq 2r + 1$, we have : $F_{r,n}(\rho) = \lfloor \frac{n}{2} \rfloor$.*

Démonstration. For the lower bound, use the paths P_n, $n \geq 2r + 1$, or the cycles $\mathcal{C}_n$, $n \geq 2r + 2$. As an upper bound, $\lfloor \frac{n}{2} \rfloor$ is valid for all connected graphs with n vertices, whether they are twin-free or not. ☐

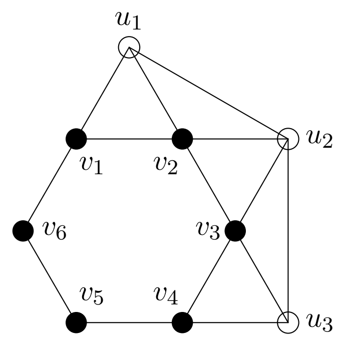


Figure A.IV – *The construction (Case 1) for $r = 2$ and $n = 9$.*

A.6 The diameter, δ

In any connected graph with n vertices, the diameter obviously lies between 1 (clique) and $n - 1$ (paths P_n).

The exact value of $f_r(\delta)$ is easy to obtain.

Theorem A.33. *For all $r \geq 1$, we have : $f_r(\delta) = r + 1$.*

Démonstration. For the lower bound, observe that if a graph G has diameter r, this means that all vertices are within distance r from each other, and condition (A.1) cannot be satisfied : G is not r-twin-free. For the upper bound, consider the cycle $\mathcal{C}_{2r+2}$, which is r-twin-free and has diameter $r + 1$. □

We are also able to give the exact values for $f_{r,n}(\delta)$ and $F_{r,n}(\delta)$.

Theorem A.34. *(a) For all $r \geq 1$, we have : $f_{r,2r+1}(\delta) = 2r$.*

(b) For all $r \geq 1$ and $n \geq 2r + 2$, we have :

$$f_{r,n}(\delta) = r + 1. \tag{A.23}$$

Démonstration. (a) When $n = 2r+1$, the graph is necessarily the path P_{2r+1}, the diameter of which is $n - 1 = 2r$.

(b) For the lower bound, see Theorem A.33. We now turn to the upper bound.

For $r = 1$ and $n \geq 3$, we can simply use the star with n vertices, so assume from now on that $r \geq 2$ and $n \geq 2r + 2$. Note that Cases 1 and 2 given below are overlapping.

Case 1 : $2r + 2 \leq n \leq 4r + 3$. Denote $k = n - (2r + 2)$; then $k < 2r + 2$. Take first the cycle $\mathcal{C}_{2r+2}$, with vertices v_i, $1 \leq i \leq 2r + 2$. Add k more vertices $u_1, u_2, \ldots, u_k$, and for $i = 1, \ldots, k$, add the edges $\{u_i, v_i\}$ and $\{u_i, v_{i+1}\}$; furthermore, add the edges $\{u_i, u_{i+1}\}$ for $i = 1, \ldots, k - 1$. Figure A.IV illustrates the case $r = 2$, $n = 9$, $k = 3$.

We claim that this graph G with n vertices is r-twin-free and has diameter $r + 1$.

To prove that G is r-twin-free, we show that for each vertex $v \in G$, the intersection $B_r(v) \cap \mathcal{C}_{2r+2}$ is unique. Indeed, for every $v_i \in \mathcal{C}_{2r+2}$ this intersection consists of all the vertices in $\mathcal{C}_{2r+2}$ except v_{i+r+1} (we always consider indices modulo $2r + 2$) and for the vertices u_j, the intersection consists of all the vertices of $\mathcal{C}_{2r+2}$ except $\{v_{j+r+1}, v_{j+r+2}\}$.

We prove that the diameter of G is $r + 1$. We have already seen that it is at least $r + 1$. Consider first a vertex $v_i \in \mathcal{C}_{2r+2}$. Every vertex in $\mathcal{C}_{2r+2}$ is trivially within distance $r + 1$. Take any u_j. Since u_j is adjacent with v_j and v_{j+1}, and at most one of these two vertices can be at distance $r + 1$ from v_i, while the other must be within distance r, we see that u_j is within distance $r + 1$ from v_i. It therefore suffices to prove that $d(u_i, u_j) \leq r + 1$ for all $i < j$. This is clear, if $j - i \leq r + 1$, because of the edges $\{u_i, u_{i+1}\}$, $\{u_{i+2}, u_{i+3}\}$,

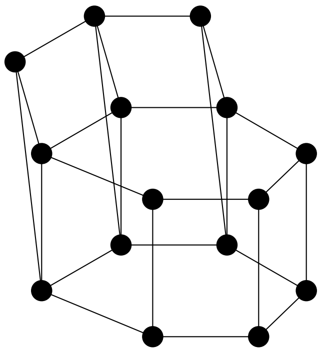


Figure A.V – *The construction (Case 2) for $r = 3$ and $n = 15$.*

..., $\{u_{j-1}, u_j\}$. Consider the case $j - i > r + 1$. Then u_j is adjacent with v_{j+1} and u_i is adjacent with v_i, and $(j+1) - i \geq r + 3$. Hence $d(v_{j+1}, v_i) \leq 2r + 2 - (r + 3) = r - 1$, and so $d(u_i, u_j) \leq r + 1$ also in this case.

Case 2 : $n \geq 4r$. Consider first the product $K_s \square \mathcal{C}_{2r}$ of K_s and $\mathcal{C}_{2r}$, where $s \geq 2$: in other words, we take s cycles

$$C(i) = \{v(i,1), v(i,2), \ldots, v(i,2r)\}, \text{ for } i = 1, 2 \ldots, s,$$

where the second index is considered modulo $2r$, and glue these together by adding the edges $\{v(i,k), v(j,k)\}$ for all k and $i \neq j$. Compare with Figure A.V : one can think of this graph as an example for $s = 3$ where three vertices from the top cycle have been deleted.

The resulting graph $K_s \square \mathcal{C}_{2r}$ is r-twin-free. Indeed, if we look at any vertex $v = v(i,t)$, then $B_r(v)$ contains all the vertices of the cycle $C(i)$ and all the vertices of $C(j)$ except $v(j, t+r)$ — the vertex diametrically opposite to $v(j,t)$ — when $j \neq i$. So, there is exactly one cycle $C(i)$ whose vertices are all in $B_r(v)$ and this tells us i, and then by looking at any other cycle $C(j)$ and determining which vertex of $C(j)$ is not in $B_r(v)$, we find t.

Clearly, the diameter of $K_s \square \mathcal{C}_{2r}$ is $r + 1$.

Consider now the general case, where $n = s \cdot 2r + k$ for some $s \geq 2$ and $0 \leq k < 2r$. It suffices to consider the case $0 < k < 2r$.

In this case we take the product $K_{s+1} \square \mathcal{C}_{2r}$ and delete the $2r - k$ consecutive vertices $v(s+1, k+1)$, ..., $v(s+1, 2r)$ from the top-most cycle $C(s+1)$. The case when $r = 3$, $n = 15$ (when $s = 2$ and $k = 3$) is illustrated in Figure A.V.

We claim that this graph is r-twin-free and has diameter $r + 1$.

We first show that the graph is r-twin-free by again showing that we can uniquely identify v if we know $B_r(v)$.

Assume that $v = v(i,t)$. If $i \leq s$, then $B_r(v)$ contains exactly one of the full cycles $C(1)$, $C(2)$, ..., $C(s)$, namely $C(i)$, the one containing v ; if $i = s + 1$, then $B_r(v)$ does not contain any of the cycles $C(1)$, $C(2)$, ..., $C(s)$. So, we can determine i.

In both cases, look at any cycle $C(j)$ not containing v. Then $B_r(v)$ contains every vertex of $C(j)$ except the one diametrically opposite to $v(j,t)$, and so we can determine t.

It remains to prove that the graph has diameter $r + 1$. Assume first that $v \in C(i)$ for some $1 \leq i \leq s$. Clearly we can reach every other vertex by taking at most $r + 1$ steps along the edges : to reach $v(j,t)$, at most r steps take us from v to $v(i,t)$, and then we just take the one step to $v(j,t)$ if $j \neq i$. Consider finally two vertices u and v of the top layer. If the distance between u and v in the original full cycle $C(s+1)$ (of which we removed some vertices) was $r - 1$ or less, we can simply move one layer down from u, use some at most $r - 1$ edges of $C(s)$, and move back up to v. If the distance between u and v in the original $C(s+1)$ was r, then the fact that we removed a set of consecutive vertices from

$C(s+1)$ guarantees that all the vertices of exactly one of the half-cycles from u to v are still all in our graph, and we are done. □

Remark A.35. *The reader could check that the construction used in Theorem A.22 can also give a graph with diameter $r+1$, but only when $\lfloor \frac{n-2}{r} \rfloor \geq 2r+2$, which is not as good a condition on n as that in Theorem A.34.*

Corollary A.36. *For all $n \geq 3$, we have : $f_{1,n}(\delta) = 2$.*

Corollary A.37. *We have : $f_{2,5}(\delta) = 4$ and for all $n \geq 6$, $f_{2,n}(\delta) = 3$.*

Démonstration. When $n = 5$, the graph is the path P_5, cf. Theorem A.34(a). When $n \geq 6$, apply (A.23). □

Finally, for $F_{r,n}(\delta)$, the obvious upper bound $n-1$ is the exact value.

Theorem A.38. *For all $r \geq 1$ and $n \geq 2r+1$, we have : $F_{r,n}(\delta) = n-1$.*

Démonstration. Again, we make use of the paths P_n, $n \geq 2r+1$. □

A.7 Recapitulatory

The tables below give a summary of the results obtained in the previous five sections, for $r = 1$, $r = 2$, and $r \geq 3$. The specific conditions on n or r are not always given. Note that we have all the exact values when $r = 1$.

$r = 1$	$f_1(.)$	$f_{1,n}(.)$	$F_{1,n}(.)$
edge number ε	2 (Pr. A.3)	$n-1$ (Pr. A.4)	$n(n-1)/2 - \lfloor n/2 \rfloor$ (Cor. A.6)
min. degree $\Delta_{\min}$	1 (Th. A.16)	1 (Th. A.16)	$n-2$ (Th. A.17)
max. stable set α	2 (Th. A.23)	2 (Th. A.24)	$n-1$ (Th. A.28)
radius ρ	1 (Th. A.30)	1 (Th. A.31)	$\lfloor \frac{n}{2} \rfloor$ (Th. A.32)
diameter δ	2 (Th. A.33)	2 (Cor. A.36)	$n-1$ (Th. A.38)

$r = 2$	$f_2(.)$	$f_{2,n}(.)$	$F_{2,n}(.)$
ε	4 (Pr. A.3)	$n-1$ (Pr. A.4)	$\gtrsim n^2/2 - 2n\log_2 n$ (ineq. (A.15)) $\lesssim n^2/2 - (n\log_2 n)/2$ (ineq. (A.15))
$\Delta_{\min}$	1 (Th. A.16)	1 (Th. A.16)	$\lfloor (n-2)/2 \rfloor$ (Th. A.18)
α	3 (Th. A.23)	3 or 4 (Cor. A.27)	$\leq n-2$ (Th. A.29) $\geq \lceil \frac{n}{2} \rceil, k + \lceil \log_2 k \rceil$, with k largest s.t. $k + 2\lceil \log_2 k \rceil \leq n-1$ (Th. A.29)
ρ	2 (Th. A.30)	2 (Th. A.31)	$\lfloor \frac{n}{2} \rfloor$ (Th. A.32)
δ	3 (Th. A.33)	3, for $n \geq 6$ (Cor. A.37)	$n-1$ (Th. A.38)

$r \geq 3$	$f_r(.)$	$f_{r,n}(.)$	$F_{r,n}(.)$
ε	$2r$ (Pr. A.3)	$n-1$ (Pr. A.4)	$\gtrsim n^2/2 - rn\log_2 n$ (ineq. (A.16)), r small w.r.t. n $\lesssim n^2/2 - .63(r - .915)n\log_2 n$ (ineq. (A.16))
$\Delta_{\min}$	1 (Th. A.16)	1 (Th. A.16)	$\geq k$ or $k-1$, with $k = \lfloor \frac{n-2}{r} \rfloor$ (Th. A.22) $\leq \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1, \frac{3n-r+2}{2(r-5)}$ (Th. A.22)
α	$r+1$ (Th. A.23)	$\leq \lfloor \frac{2n}{k+2} \rfloor$ (k even), $\leq \lfloor \frac{2n}{k+1} \rfloor$ (k odd), with $k = \lfloor \frac{n-2}{r} \rfloor$ (Th. A.25) $\geq r+1$ (Th. A.25)	$\geq \lceil \frac{n}{2} \rceil, k + \lceil \log_2 k \rceil \lfloor \frac{r}{2} \rfloor$, with k largest s.t. $k + r\lceil \log_2 k \rceil \leq n-1$ (Th. A.29) $\leq n-r$ (Th. A.29)
ρ	r (Th. A.30)	r (Th. A.31)	$\lfloor \frac{n}{2} \rfloor$ (Th. A.32)
δ	$r+1$ (Th. A.33)	$r+1$ (Th. A.34)	$n-1$ (Th. A.38)

Annexe B

Induced Paths in Twin-Free Graphs

David Auger [1]
david.auger@telecom-paristech.fr

Abstract

Let $G = (V, E)$ be a simple, undirected graph. Given an integer $r \geq 1$, we say that G is *r-twin-free* (or *r-identifiable*) if the balls $B(v, r)$ for $v \in V$ are all different, where $B(v, r)$ denotes the set of all vertices which can be linked to v by a path with at most r edges. These graphs are precisely the ones which admit r-identifying codes. We show that if a graph G is r-twin-free, then it contains a path on $2r + 1$ vertices as an induced sugbraph, i.e. a chordless path.

Keywords : graph theory ; identifying codes ; twin-free graphs ; induced path ; radius.
2000 *Mathematics Subject Classification :* 05C12, 05C99.

B.1 Notation and definitions

Let $G = (V, E)$ be a simple, undirected graph. We will denote an edge $\{x, y\} \in E$ simply by xy. A *path* in G is a sequence $P = v_0v_1 \cdots v_k$ of vertices such that for all $0 \leq i \leq k-1$ we have $v_iv_{i+1} \in E$; if $v_0 = x$ and $v_k = y$, we say that P is a path *between* x and y.

The *length* of a path $P = v_0v_1 \cdots v_k$ is the number of edges between consecutive vertices, i.e. k. If $x, y \in V$, we define the distance $d(x, y)$ to be the minimum length of a path between x and y. Then a *shortest path* between x and y is a path between x and y of length precisely $d(x, y)$. If $r \geq 0$, $B(x, r)$ will denote the *ball* of centre x and radius r, which is the set of all vertices v of G such that $d(x, v) \leq r$.

If $P = v_0 \cdots v_k$ is a path in G, a *chord* in P is any edge $v_iv_j \in E$ with $|i - j| \neq 1$. A path is *chordless* if it has no chord ; in this case there is an edge between two vertices of the

1. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

path v_i and v_j if and only if i and j are consecutive, i.e. $|i-j|=1$. It is straightforward to see that any shortest path is chordless.

If $x \in V$, we define the *excentricity* of x by

$$\text{exc}(x) = \max_{v \in V} d(x,v).$$

The *diameter* of G is the maximum excentricity of a vertex in G, whereas the *radius* rad(G) of G is the minimum excentricity of a vertex in G. A vertex x such that $\text{exc}(x) = \text{rad}(G)$ is a *centre* of G. So G has radius $t \geq 1$ and x is a centre of G if and only if $B(x,t) = V$ whereas $B(v,t-1) \neq V$ for all $v \in V$.

If $W \subset V$, the *sugbraph* of G *induced* by W is the graph whose set of vertices is W and whose edges are all the edges $xy \in E$ such that x and y are in W. We denote this graph by $G[W]$; if $W = V \setminus \{v\}$, we simply write $G[V-v]$. An *induced path* in G is a subset P of V such that $G[P]$ is a path ; equivalently, the vertices in P define a chordless path in G. All these terminology and notation being standard, we refer to [43] for further explanation.

Two distinct vertices x and y are called r*-twins* if $B(x,r) = B(y,r)$. If there are no r-twins in G, we say that G is r*-twin-free.*

B.2 Motivations and main results

The notion of identifying code in a graph was introduced by Karpovsky, Chakrabarty and Levitin in [65]. For $r \geq 1$, an r-identifying code in $G = (V,E)$ is a subset $\mathcal{C}$ of V such that the sets

$$I_{\mathcal{C}}(v) = B(v,r) \cap \mathcal{C} \text{ for } v \in V$$

are all distinct and non-empty. The original motivation for identifying codes was the fault diagnosis in multiprocessor systems ; we refer to [4], [65] or [74] for further explanation and applications. The interested reader can also find a nearly exhaustive bibliography in [73].

Given a graph $G = (V,E)$, it is easily seen that there exists an r-identifying code in G if and only if V itself is an r-identifying code, which precisely means that G is r-twin-free. Different structural properties which are worth investigating arise when considering a connected r-twin-free graph with $r \geq 1$. For instance, it has been proved in [29] that an r-twin-free graph always contains a path, not necessarily induced, on $2r+1$ vertices. In the same article, the authors conjectured that we can always find such a path as an induced subgraph of G. We prove this conjecture as a corollary from Theorem B.1.

Let us denote by $p(G)$ the maximum number of vertices of an induced path in G. We prove the following theorem and corollary, which we formulate for connected graphs without loss of generality.

Theorem B.1. *Let $G = (V,E)$ be a connected graph with at least two vertices, and with a centre $c \in V$ such that no neighbour of c is a centre. Then*

$$p(G) \geq 2 \; rad(G) + 1.$$

This implies :

Corollary B.2. *Let G be a connected graph with at least two vertices, and $r \geq 1$. If G is r-twin-free then*

$$p(G) \geq 2r + 1.$$

B.3 Proof of the theorem

A different proof for Corollary B.2 can be found in [4]. The one we present here is much shorter and is based on the article by Erdös, Saks and Sós [46] where the following theorem can be found. The authors give credit to Fan Chung for the proof.

Theorem B.3. *(Chung) For every connected graph $G = (V, E)$ we have*

$$p(G) \geq 2\ rad(G) - 1.$$

We require the following lemma, inspired by [46], in order to prove Theorem B.1.

Lemma B.4. *Let $t \geq 2$ and $G = (V, E)$ be a graph such that there are in G two vertices v_0 and v_t with $d(v_0, v_t) = t$, a shortest path $v_0v_1v_2 \cdots v_t$ between v_0 and v_t, and a vertex w such that $d(v_0, w) \leq t-1$ and $d(v_2, w) \geq t$ (see fig. B.I). Then there exists an induced path on $2t-1$ vertices in G.*

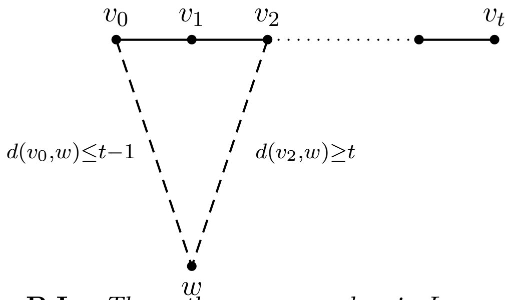


Figure B.I – *The path $v_0 \cdots v_t$ and w in Lemma B.4.*

Proof. In the case $t = 2$, the shortest path $v_0v_1v_2$ itself is an induced path on $2t-1 = 3$ vertices ; so we suppose now that $t \geq 3$. First observe that since $d(v_2, w) \geq t$ we have $w \neq v_i$ for all $i \in \{0, 1, \cdots, t\}$. Consider a shortest path P between v_0 and w, and let $u \in P$, distinct from v_0. Let $i \geq 2$; we show that $d(u, v_i) \geq 2$. First we have

$$d(v_0, v_i) = i \leq d(v_0, u) + d(u, v_i)$$

and second

$$t \leq d(v_2, w) \leq d(v_2, v_i) + d(v_i, u) + d(u, w)$$

with $d(v_2, v_i) = i - 2$ because $i \geq 2$. Summing these two inequalities we get

$$t + i \leq d(v_0, u) + d(u, w) + 2d(v_i, u) + i - 2$$

and since

$$d(v_0, u) + d(u, w) = d(v_0, w)$$

we deduce

$$t + 2 \leq d(v_0, w) + 2d(u, v_i).$$

But we have $d(v_0, w) \leq t - 1$ and so

$$d(u, v_i) \geq \frac{3}{2}.$$

Let us note that since $d(v_2, w) \geq t$, we have $d(v_0, w) \geq t - 2$ and so P consists of v_0 and at least $t - 2 \geq 1$ other vertices, i.e. at least $t - 1$ vertices. We proved that u satisfies $d(u, v_i) \geq 2$ for $i \geq 2$, so u is distinct from all the v_i's and furthermore can be adjacent only to v_1 or v_0 (see fig. B.II).

Now consider two cases :

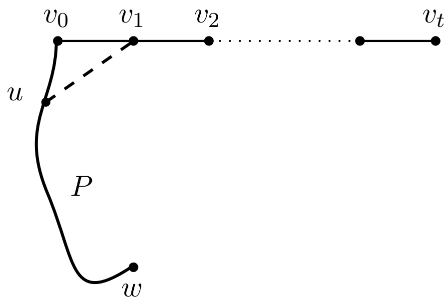


Figure B.II – *The vertex u can only be adjacent to v_1 or v_0 in Lemma B.4.*

- if no vertex $u \in P \setminus \{v_0\}$ is adjacent to v_1, then P extended by $v_1 \cdots v_t$ is an induced path of G on at least $(t-1)+t = 2t-1$ vertices ;
- if there is a vertex $u \in P \setminus \{v_0\}$ adjacent to v_1, then

$$t \leq d(v_2, w) \leq d(v_2, v_1) + d(v_1, u) + d(u, w)$$

and so $d(u, w) \geq t-2$. Since we have $d(v_0, u) + d(u, w) = d(v_0, w) \leq t-1$, it follows that we must have $d(v_0, u) = 1$ and $d(u, w) = t-2$. The path $w \cdots u v_1 \cdots v_t$ is then an induced path of G on $2t-1$ vertices. □

For sake of completeness, we rephrase the end of the proof of Theorem B.3 in [46]. Consider a connected graph G of radius $t \geq 1$; if $t = 1$, then the result is trivial. Suppose now that $t \geq 2$; we show that the vertices $v_0, v_1, \cdots v_t$ and w as in Lemma B.4 exist. To see this, consider the collection of connected induced subgraphs H of G whose radius is at least t, and choose one with the smallest possible number of vertices. Let V_H be the vertex-set of H.

There exists in H a vertex v_t which is not a cutvertex ; by minimality of H, the connected induced subgraph $H[V_H - v_t]$ of H must have radius at most $t-1$. If we consider a centre v_0 of $H[V_H - v_t]$, we must have $d(v_0, w) \leq t-1$ for all the vertices $w \neq v_t$ in H ; but since H has radius at least t we also have $d(v_0, v_t) = t$. Let $v_0 v_1 v_2 \cdots v_t$ be a shortest path between v_0 and v_t. Since H has radius t, there exists a vertex w such that $d(v_2, w) \geq t$, and we have $d(v_0, w) \leq t-1$ because w cannot be v_t. So we can choose this w and apply Lemma B.4.

Proof of Theorem B.1. Let $G = (V, E)$ be a graph of radius $t \geq 1$ with a centre $c \in V$ such that no neighbour of c is a centre. We will apply Lemma B.4 with $t+1$ instead of t ; to do this, we have to find vertices $v_0, v_1, \cdots v_{t+1}$ and w ; so let us denote the center c by v_1. We define $N(v_1)$ to be the set of neighbours of v_1. We can choose a vertex v_0 in $N(v_1)$ such that $B(v_0, t)$ is not strictly contained in another $B(x, t)$ for $x \in N(v_1)$: take for instance $v_0 \in N(v_1)$ such that $B(v_0, t)$ is of maximal cardinality. Since v_0 is not a centre, there exists a vertex $v_{t+1} \in V$ such that $d(v_0, v_{t+1}) = t+1$. Then we must have $d(v_1, v_{t+1}) \geq t$, and so $d(v_1, v_{t+1}) = t$ because v_1 is a centre. Consider a shortest path $v_1 v_2 \cdots v_{t+1}$ between v_1 and v_{t+1} ; then $v_0 v_1 v_2 \cdots v_{t+1}$ is a shortest path between v_0 and v_{t+1}. Now, if we show that there exists a vertex w such that $d(v_2, w) \geq t+1$ and $d(v_0, w) \leq t$, we can apply Lemma B.4. So, assume that such a vertex w does not exist : this means that all the vertices w with $d(v_0, w) \leq t$ must satisfy $d(v_2, w) \leq t$, and so $B(v_0, t) \subset B(v_2, t)$. By maximality of $B(v_0, t)$, we must then have $B(v_0, t) = B(v_2, t)$; but this is impossible, since we have $v_{t+1} \in B(v_2, t) \setminus B(v_0, t)$. This contradiction shows that we can apply Lemma B.4, and so there exists in G an induced path on $2(t+1)-1=2t+1$ vertices ; thus we have

$$p(G) \geq 2\text{rad}(G)+1.$$

□

Proof of Corollary B.2. Let G be a graph, x a center of G and y a neighbour of x. Then by definition $B(x, \text{rad}(G)) = V$, and for all $z \in V$ we have

$$d(y,z) \leq d(z,x) + d(x,y) \leq \text{rad}(G)+1.$$

So

$$B(x,r) = B(y,r) = V$$

for all $r \geq \text{rad}(G)+1$. Suppose now that G is r-twin-free ; then we must have $\text{rad}(G) \geq r$.

Now, either $\text{rad}(G) \geq r+1$ and we can apply Theorem B.3, or $\text{rad}(G) = r$. But in the latter case, centers are r-twins so there can only be one in G ; in particular we can apply Theorem B.1 and so

$$p(G) \geq 2\text{rad}(G)+1 = 2r+1.$$

□

B.4 Conclusion and perspectives

For $n \geq 1$, we denote by P_n the path on n vertices, i.e. the graph consisting of n vertices $v_0, v_1, \cdots, v_{n-1}$ and the $n-1$ edges $v_i v_{i+1}$ for $0 \leq i \leq n-1$. As the path P_{2r+1} on $2r+1$ vertices is itself r-twin-free, the previous results show that P_{2r+1} is the only minimal r-twin-free graph for the induced subgraph relationship. Indeed, we have :

An r-twin-free graph contains a path P_{2r+1} as an induced sugbraph, and P_{2r+1} is r-twin-free.

One could wonder how these results could be extended to different cases. For instance, we have :

An r-twin-free and 2-connected graph G contains a cycle with at least $2r+2$ vertices as a subgraph ; and the cycle C_k on k vertices is r-twin-free if and only if $k \geq 2r+2$ (and is, of course, 2-connected).

Let us recall that a graph G is 2-connected if and only if for every pair (x, y) of distinct vertices, there exist at least two paths P_1 and P_2 between x and y in G, such that there are no common vertices to P_1 and P_2 except x and y (see [43], pp. 55-57 for more details). Since an r-twin-free graph has a diameter at least $r+1$, the result above easily follows. This shows that the cycles C_k with $k \geq 2r+2$ are the minimal graphs for the subgraph relationship in the class of 2-connected, r-twin-free graphs. But in this case, the result cannot be extended to the induced subgraph relationship. Indeed, for $r \geq 1$ consider the Cartesian product of a path P_{2r+1} with K_2 (see fig. B.III). One can check that this graph is 2-connected, r-twin-free and does not contain a cycle with more than $2r+2$ vertices as an induced subgraph. For $r=1$, see the counterexample on fig. B.IV

As a conclusion, we leave open the same problem in the class of k-connected graphs with $k \geq 3$:

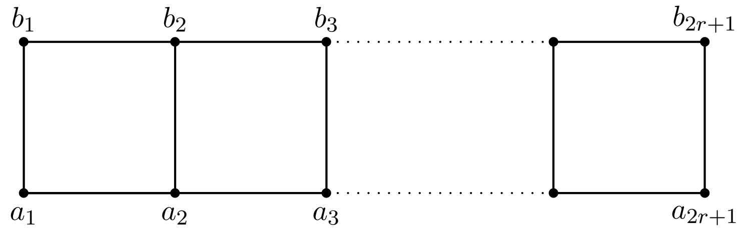


Figure B.III – *A 2-connected, r-twin-free graph which does not contain a cycle C_k with $k \geq 2r+2$ as an induced subgraph ($r \geq 2$).*

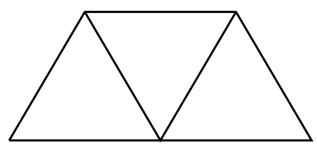


Figure B.IV – *A 2-connected, 1-twin-free graph which does not contain a cycle C_k with $k \geq 4$ as an induced subgraph.*

what are the minimal elements of the class of 3-connected, r-twin-free graphs, for the subgraph relationship, or the induced subgraph relationship ?

A first step would be to determine the smallest cardinality for a k-connected r-twin-free graph.

Annexe C

Existence d'un cycle de longueur au moins 7 dans un graphe sans $(1, \leq 2)$-jumeaux

David Auger [1], **Irène Charon** [1],
Olivier Hudry [1], **Antoine Lobstein** [2]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Résumé

On considère un graphe G simple non orienté. On appelle boule d'un sous-ensemble Y de sommets de G l'ensemble des sommets de G à distance au plus 1 d'un sommet de Y. On suppose que les boules des sous-ensembles des sommets de G de cardinal au plus 2 sont toutes distinctes. On montre qu'alors G possède un cycle de longueur au moins 7.

Mots-clés : graphe non orienté, sous-ensembles jumeaux, graphe identifiable, code identifiant, cycle de longueur maximum.

C.1 Introduction

On s'intéresse à un graphe non orienté, simple, fini $G = (X, E)$, où X désigne l'ensemble des sommets de G et E l'ensemble de ses arêtes.

Si r est un entier positif et x un sommet de G, on nomme *boule de x de rayon r*, ou *r-boule de x*, et on note $B_r(x)$ l'ensemble constitué des sommets de G à distance au plus r de x. Si Y est une partie de X, on définit la *boule de Y de rayon r*, ou *r-boule de Y*, et on note $B_r(Y)$ l'ensemble :

$$B_r(Y) = \bigcup_{y \in Y} B_r(y).$$

2. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

Pour tout sommet x, on pose $B(x) = B_1(x)$ et on appelle cet ensemble *boule de x* : en d'autres termes, la boule de x est constituée de x et de ses voisins ; pour toute partie Y de X, on pose $B(Y) = B_1(Y)$ et on appelle cet ensemble *boule de Y*.

Deux parties de X sont dites *séparées* si leurs r-boules sont distinctes. Pour un entier $\ell \geq 1$, le graphe G est dit *sans $(r, \leq \ell)$-jumeaux* si deux quelconques sous-ensembles distincts de X de cardinal au plus ℓ sont séparés. Dans un graphe sans $(r, \leq \ell)$-jumeaux, pour toute partie V de X, il existe au plus un sous-ensemble Y de X, de cardinal au plus ℓ, pour lequel $B_r(Y) = V$: les sous-ensembles de X de cardinal au plus ℓ sont caractérisés par leur r-boule. On dit aussi dans ce cas que G est *$(r, \leq \ell)$-identifiable*, ou qu'il admet un *code $(r, \leq \ell)$-identifiant*. Voir, parmi d'autres, les références [53] [61] [69] [70] [74]].

Nous nous restreindrons ici au cas $r = 1$, $\ell = 2$. On sait peu de choses sur la structure des graphes sans $(1, \leq 2)$-jumeaux ; l'objectif de cet article est de prouver qu'un graphe non orienté, connexe, d'ordre au moins 2 sans $(1, \leq 2)$-jumeaux possède un cycle *élémentaire* (ne passant pas deux fois par le même sommet) de longueur au moins 7.

Nous rappelons ici quelques définitions classiques pour un graphe $G = (X, E)$ [16],[43]. Etant donné un sous-ensemble $X' \subseteq X$ de sommets de G, le *sous-graphe de G induit ou engendré par X'* est le graphe $G' = (X', E')$ où

$$E' = \{\{u, v\} \in E : u \in X', v \in X'\}.$$

Un *sous-graphe partiel* de G est un graphe $G'' = (X'', E'')$, avec $X'' \subseteq X$ et

$$E'' \subseteq \{\{u, v\} \in E : u \in X'', v \in X''\}.$$

Un *point* (ou *sommet*) *d'articulation* de G est un sommet u de X tel que le sous-graphe induit par $X \setminus \{u\}$ voit son nombre de composantes connexes augmenter par rapport à G. Un *isthme* de G est une arête a de E telle que le sous-graphe partiel $(X, E \setminus \{a\})$ voit son nombre de composantes connexes augmenter par rapport à G. Lorsque G est connexe, la suppression d'un point d'articulation ou d'un isthme déconnecte le graphe. Plus généralement, un graphe *h-connexe*, $h \geq 1$, est un graphe G pour lequel le nombre minimum de sommets à supprimer pour déconnecter G, ou pour le réduire à un singleton, est au moins h. Une *composante h-connexe* de G est un sous-graphe induit h-connexe et maximal, pour l'inclusion, dans G.

On appelle *bloc* de G tout sous-graphe induit maximal sans sommet d'articulation, et on appelle *pont* un sous-graphe induit constitué de deux sommets adjacents, reliés par une arête qui est un isthme de G.

Enfin on utilisera la notation $\mathcal{C}_i$ (respectivement, $\mathcal{C}_{\geq i}$) pour un cycle de longueur i (respectivement, au moins i), $i \geq 3$.

Dans tout cet article, les chaînes ou les cycles considérés seront élémentaires, et $G = (X, E)$ sera un graphe non orienté, simple, d'ordre au moins 2, et qu'on supposera connexe : si G n'était pas connexe, le résultat serait obtenu en choisissant une composante connexe de G ayant au moins 2 sommets.

C.2 Choix d'un *bloc feuille* de G

Les blocs de G sont des composantes 2-connexes ou des ponts. Le graphe représenté à gauche de la figure C.I possède cinq blocs : $\{a, b, c, d\}$, $\{c, e\}$, $\{g, h, i\}$, $\{e, f, g\}$, et $\{f, j\}$, qui sont entourés en pointillé. Deux blocs de G soit ont une intersection vide, soit s'intersectent en un sommet d'articulation de G. Considérons le graphe G' dont les sommets

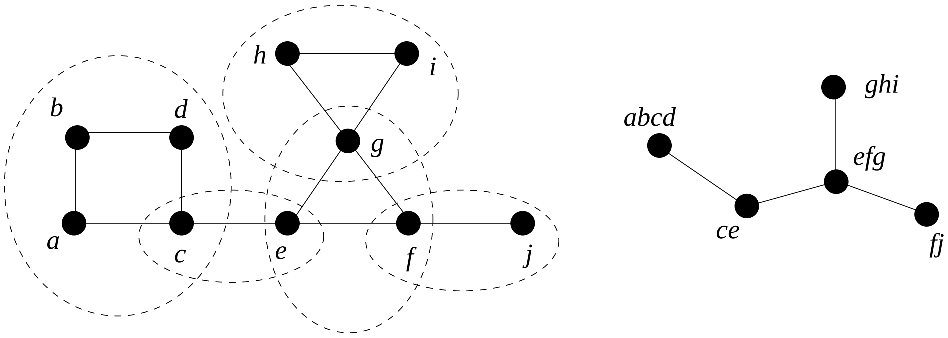


Figure C.I – *Exemple de graphes G et G'.*

correspondent aux blocs de G, et tel que deux sommets sont adjacents s'ils correspondent à deux blocs ayant un sommet en commun : le graphe G' est un arbre. Nous appelons *bloc feuille* de G un bloc de G correspondant à une feuille de G'. Le graphe représenté à gauche de la figure C.I possède trois blocs feuilles.

On introduit la définition suivante :

Définition C.1. *Soient $G = (X, E)$ un graphe non orienté, $Y \subset X$ un sous-ensemble de sommets, y un sommet dans Y, et s un sommet de $X \setminus Y$. On appelle (G, s, Y, y)-chaîne une chaîne de G dont une extrémité est s, l'autre extrémité, t, appartient à $Y \setminus \{y\}$, et dont les sommets autres que t sont dans $X \setminus Y$.*

On utilisera de manière répétée la proposition suivante :

Proposition C.1. *Soient $G = (X, E)$ un graphe non orienté connexe, H une composante 2-connexe de G, Y un sous-ensemble d'au moins 2 sommets dans H, y un sommet de Y qui n'est pas sommet d'articulation de G, et s un voisin de y qui n'appartient pas à Y. Alors, le sommet s appartient à H et il existe une (H, s, Y, y)-chaîne.*

Preuve. On note $G \setminus \{y\}$ le graphe obtenu à partir de G en retirant le sommet y. Le sommet y n'étant pas sommet d'articulation, le graphe $G \setminus \{y\}$ est connexe. Il existe donc dans $G \setminus \{y\}$ une chaîne reliant s et un sommet t de $Y \setminus \{y\}$ dont les sommets autres que t sont dans $X \setminus Y$, c'est-à-dire une (G, s, Y, y)-chaîne ; si on concatène cette chaîne avec l'arête $\{s, y\}$, on obtient une chaîne C entre y et t qui sont deux sommets distincts de la composante 2-connexe H ; l'union de H et de C est encore 2-connexe ; la maximalité de H en tant que sous-graphe 2-connexe entraîne que la chaîne C est une chaîne dans H. □

La proposition C.1 exprime que, si on cherche à << sortir >> d'un sous-ensemble Y de 2 ou plus sommets d'une composante 2-connexe H à partir d'un non point d'articulation y, alors on reste dans H et on << revient >> dans Y, ailleurs qu'en y.

On suppose maintenant et dans toute la suite que G est sans $(1, \leq 2)$-jumeaux.

Remarquons d'abord que G ne peut pas avoir de sommet de degré 1 ; en effet, supposons qu'il existe un sommet x de degré 1 et notons y son unique voisin : l'ensemble $\{y\}$ ne serait pas séparé de l'ensemble $\{x, y\}$; ceci fait d'ailleurs partie d'un résultat plus général sur les graphes sans $(1, \leq \ell)$-jumeaux, dont le degré minimum est au moins ℓ ([70], Th. 8). En conséquence, un bloc feuille de G ne peut pas être un pont : les blocs feuilles de G sont des composantes 2-connexes. On pourra en particulier leur appliquer la proposition C.1. Nous notons H un bloc feuille de G ; le graphe H possède au moins un cycle.

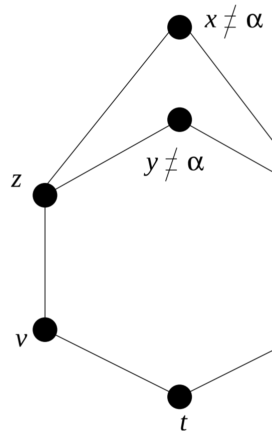


Figure C.II – *Le graphe L du lemme C.2.*

Le graphe H est soit G tout entier et dans ce cas ne possède pas de point d'articulation, soit possède un et un seul sommet, α, qui soit sommet d'articulation de G. Dans toute la suite, on garde ces notations pour le graphe (composante 2-connexe) H et le sommet α **s'il existe**.

C.3 Le plus long cycle de H n'est pas de longueur 6

Le lemme C.2 sera utilisé de manière répétée pour démontrer les lemmes C.3–C.5, lesquels stipulent que si H admet certains sous-graphes partiels, alors, sous certaines conditions, H admet un $\mathcal{C}_{\geq 7}$ comme sous-graphe partiel. Le lemme C.6 conclut cette section, de loin la plus longue de cet article, en établissant que le plus long cycle de H n'est pas de longueur 6.

Lemme C.2. *On suppose que le plus long cycle de H est de longueur 6 ; si H possède le graphe L représenté sur la figure C.II comme sous-graphe partiel, avec $x \neq \alpha$ et $y \neq \alpha$, alors le sommet t est adjacent à x ou y mais pas aux deux, et les sommets x et y n'ont pas d'autres voisins dans G que z, u, et, pour exactement l'un d'entre eux, t.*

Preuve. On suppose que H n'admet pas de cycle de longueur au moins 7 et qu'il possède le graphe L comme sous-graphe partiel, avec $x \neq \alpha$ et $y \neq \alpha$. Notons Y l'ensemble des 7 sommets de L.

On montre d'abord que les voisins dans G de x et de y sont dans l'ensemble $\{z, u, t\}$. Supposons au contraire que le sommet x a un voisin s appartenant à $X \setminus \{z, u, t\}$.

Si s appartient à Y, il s'agit de y, v, ou w.

Si s n'appartient pas à Y, puisque x n'est pas le sommet d'articulation, on peut utiliser la proposition C.1 : le sommet s appartient à H et il existe une (H, s, Y, x)-chaîne.

Que le sommet s appartienne ou non à Y, il existe donc une chaîne C de longueur au moins 1 reliant x et $Y \setminus \{x\}$, autre que les arêtes $\{x, z\}$, $\{x, u\}$ et $\{x, t\}$, et dont les sommets, sauf les extrémités, n'appartiennent pas à Y ; on examine les différents cas envisageables, qui sont représentés sur la figure C.III.

- (a) Si C relie x et z, C est de longueur au moins 2 ; en la concaténant avec la chaîne z, v, t, w, u, x, on a un cycle de longueur supérieure ou égale à 7, représenté en gras en figure C.III(a) ; ce cas est impossible, de même que le cas où C relie x et u.
- (b) Si C relie x et y, cette chaîne concaténée avec la chaîne y, z, v, t, w, u, x forme un $\mathcal{C}_{\geq 7}$; ce cas est impossible.

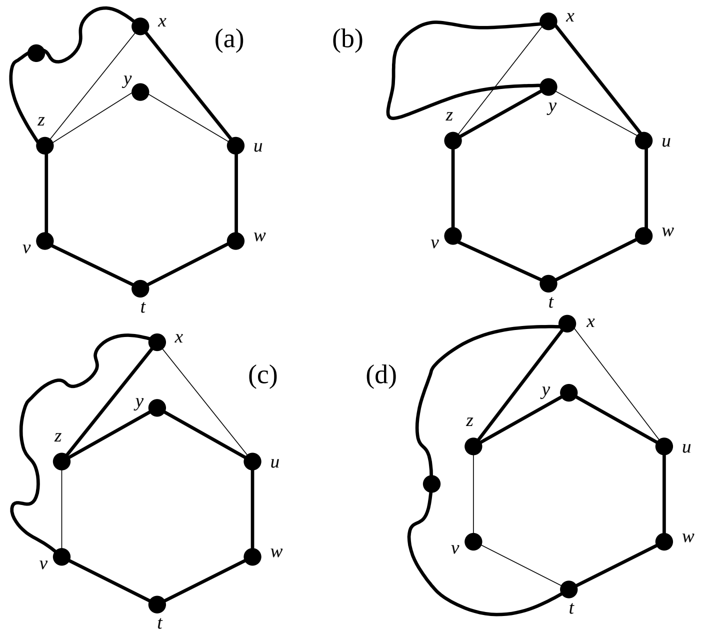


Figure C.III – *Illustrations pour la preuve du lemme C.2.*

– (c) Si C relie x et v, cette chaîne concaténée avec la chaîne v, t, w, u, y, z, x forme un $\mathcal{C}_{\geq 7}$. Il est de même impossible que C relie x et w.
– (d) Enfin, si C relie x et t, alors C est de longueur au moins 2 et en la concaténant avec la chaîne t, w, u, y, z, x, on forme un $\mathcal{C}_{\geq 7}$, encore une contradiction.

Aucun cas n'est donc possible, et les voisins de x sont dans l'ensemble $\{z, u, t\}$; il en est de même pour le sommet y.

Par ailleurs, on a : $B(\{z, x\}) \supset \{x, y, z, u\}$ et $B(\{z, y\}) \supset \{x, y, z, u\}$. Pour que les ensembles $\{z, x\}$ et $\{z, y\}$ soient séparés, il est nécessaire d'utiliser le sommet t, et donc un, et un seul, des sommets x et y est adjacent à t, ce qui achève la preuve du lemme C.2. □

Lemme C.3. *Si le graphe H admet le graphe L représenté par la figure C.II avec $x \neq \alpha$ et $y \neq \alpha$ comme sous-graphe partiel, alors le graphe H possède un cycle de longueur au moins 7 comme sous-graphe partiel.*

Preuve. On suppose que H ne possède pas de $\mathcal{C}_{\geq 7}$ et admet le graphe L avec $x \neq \alpha$ et $y \neq \alpha$ comme sous-graphe partiel. On nomme encore Y l'ensemble des 7 sommets de L.

On peut supposer que, si α n'appartient pas à Y, il n'existe pas la chaîne z, α, t ; en effet, s'il existe la chaîne z, α, t avec α hors de Y, on supprime de L la chaîne z, v, t qu'on remplace par la chaîne z, α, t et le sommet α prend le nom de v. On peut de même supposer que, si α n'appartient pas à Y, la chaîne u, α, t n'existe pas.

Si le sommet α se trouve en z ou en w, on renomme les sommets en échangeant les noms z et u ainsi que les noms v et w ; on peut donc supposer, sans perte de généralité, que le sommet α n'est ni en z, ni en w.

Le graphe L considéré dorénavant possède les propriétés suivantes :
– L correspond à la figure C.II,
– $x \neq \alpha$, $y \neq \alpha$, $z \neq \alpha$, et $w \neq \alpha$,
– s'il existe la chaîne z, α, t, alors α appartient à Y,
– s'il existe la chaîne u, α, t, alors α appartient à Y.

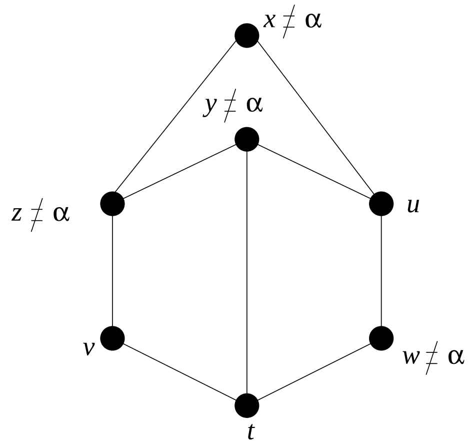


Figure C.IV – *Le graphe L avec l'arête {y,t}.*

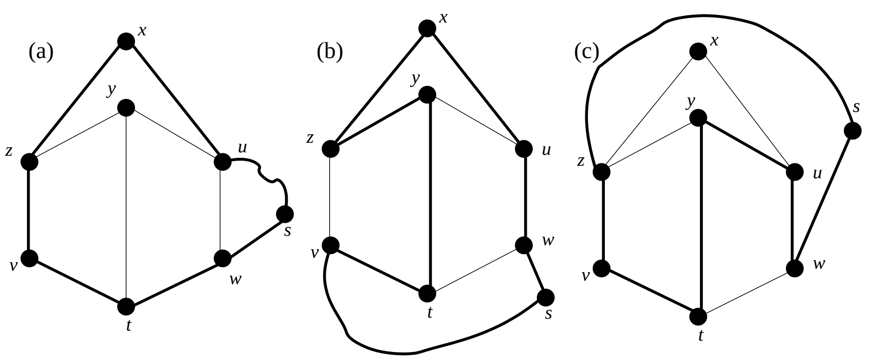


Figure C.V – *Lemme C.3, illustrations pour le résultat 1.*

On peut supposer en outre, d'après le lemme C.2, que y est adjacent à t et on sait qu'alors les sommets x et y n'ont pas d'autres voisins dans G que ceux représentés sur la figure C.IV. Le graphe H possède donc le graphe représenté sur la figure C.IV comme sous-graphe partiel.

Pour établir le lemme C.3, on procède par étapes, en établissant des résultats intermédiaires, de 1 à 7.

1. *Le sommet w n'a aucun voisin hors de Y.*

   Supposons que w ait un voisin s n'appartenant pas à Y (voir figure C.V) ; w étant différent de α, il y a une (H, s, Y, w)-chaîne C. D'après le lemme C.2, les sommets x et y ont tous leurs voisins dans Y : la chaîne C ne peut pas arriver en x ou y. La chaîne C ne peut arriver ni en u ni en t, sans quoi on aurait un $\mathcal{C}_{\geq 7}$, représenté en gras en figure C.V(a) dans le cas où la chaîne arrive en u. Si la chaîne C arrivait en v, on aurait un $\mathcal{C}_{\geq 8}$ et si elle arrivait en z, on aurait un $\mathcal{C}_{\geq 7}$: la chaîne C ne peut arriver en aucun sommet de Y. En conséquence, w n'a pas de voisin hors de Y.

2. *Si le sommet v est différent de α, alors v n'a aucun voisin hors de Y.*

   Ce résultat s'obtient exactement de la même manière que le résultat 1.

3. *Il n'existe pas de sommet n'appartenant pas à Y et différent de α qui soit adjacent à la fois à z et u.*

   En effet, supposons qu'un sommet s différent de α et hors de Y soit adjacent à la fois à z et u (voir figure C.VI) ; d'après le lemme C.2, le sommet x n'étant pas adjacent

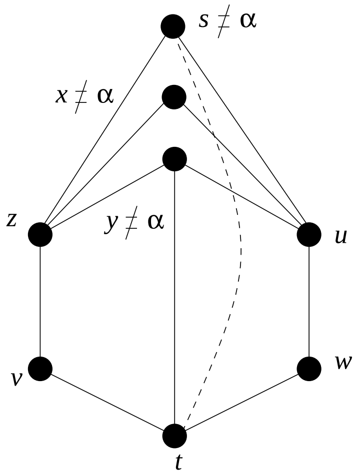


Figure C.VI – *Lemme C.3, illustration pour le résultat 3.*

à t et ni x ni s n'étant point d'articulation, le sommet s est adjacent à t. Ni s ni y n'étant point d'articulation, le fait que s et y soient adjacents à t contredit le lemme C.2.

4. *Si v est différent de α et si le sommet z a un voisin s n'appartenant pas à Y, alors s est le sommet α et on a la chaîne z, α, u.*

   On suppose que v est différent de α et on rappelle que le sommet z est différent de α ; on suppose que le sommet z a un voisin s hors de Y. Le sommet z étant différent de α, la proposition C.1 montre qu'il existe une (H, s, Y, z)-chaîne, qu'on nomme C. La chaîne C ne peut arriver ni en x, ni en y, ni en v, car on aurait alors un $\mathcal{C}_{\geq 7}$. Pour la même raison, elle ne peut pas non plus arriver en w, cf. figure C.V(c).

   Supposons maintenant que la chaîne C arrive en t ; nécessairement, C est de longueur 1 (il s'agit de l'arête $\{s, t\}$), sinon on aurait un $\mathcal{C}_{\geq 7}$; or on a choisi L de sorte que, si la chaîne z, α, t existe, alors $\alpha \in Y$: le sommet s est donc différent de α ; d'après le lemme C.2, appliqué à s et v, soit v, soit s doit être adjacent à u et les sommets s et v n'ont aucun voisin n'appartenant pas à $\{z, t, u\}$. On va montrer que v ne peut pas être adjacent à u ; pour cela, on suppose que le sommet v est adjacent au sommet u. En rappelant que le sommet y n'a aucun voisin n'appartenant pas à $\{z, u, t\}$, on a (voir figure C.VII) :

   $$B(\{t, y\}) = B(\{t, v\}) = \{y, z, t, u, v\} \cup B(t).$$

   Les paires de sommets $\{t, y\}$ et $\{t, v\}$ ne sont donc pas séparées. Le sommet v ne peut donc pas être adjacent à u. On montrerait de même que, si s était adjacent à u, les paires $\{t, y\}$ et $\{t, s\}$ ne seraient pas séparées. Ni v, ni s ne peut être adjacent à u : la chaîne C ne peut pas arriver en t.

   Il reste l'éventualité que C arrive en u. Alors, comme précédemment, on voit que C est nécessairement de longueur 1 : on a la chaîne z, s, u. Le résultat 3 montre que le sommet s est α, ce qui achève la preuve du résultat 4.

5. *Si u est différent de α et si u a un voisin s n'appartenant pas à Y, alors s est le sommet α et on a la chaîne u, α, z.*

   On suppose que u est différent de α et, par hypothèse, w est différent de α. La preuve du résultat 4 s'appuie sur la propriété que les sommets z et v sont autres que α. On peut reprendre cette preuve pour démontrer symétriquement le résultat 5.

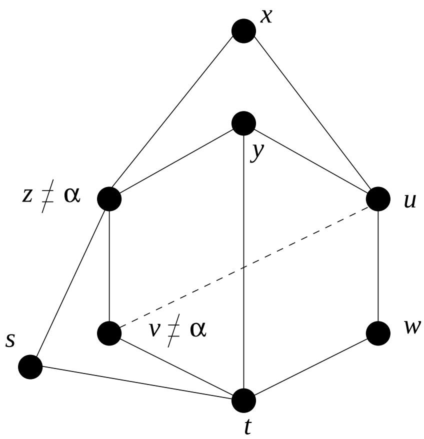


Figure C.VII – *Lemme C.3, illustration pour le résultat 4, lorsque C arrive en t.*

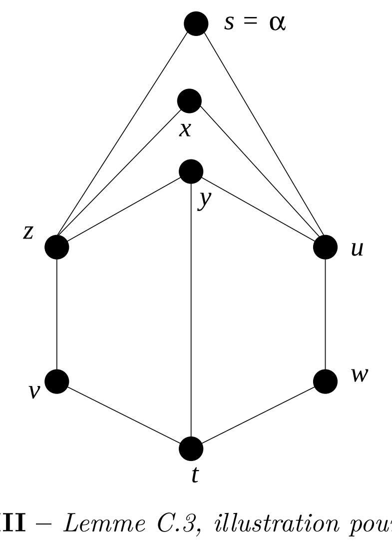


Figure C.VIII – *Lemme C.3, illustration pour le résultat 6.*

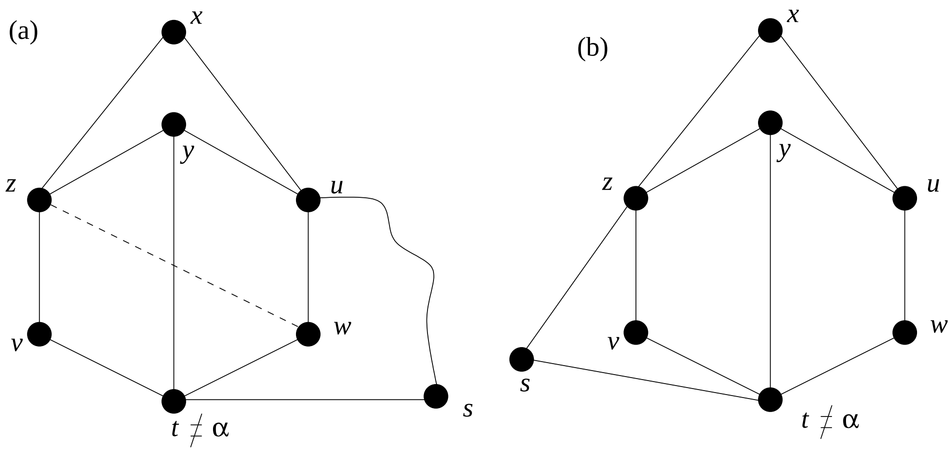


Figure C.IX – *Lemme C.3, illustrations pour le résultat 7.*

6. *α est en u ou v.*

   Supposons que u et v soient différents de α. D'après les résultats 1 et 2, les sommets v et w n'ont alors aucun voisin hors de Y ; d'après les résultats 4 et 5, les sommets z et u peuvent seulement avoir le sommet α comme voisin hors de Y, voisin qui leur est alors commun (voir figure C.VIII). On a :

   $$B(\{w,z\}) = B(\{v,u\}) = Y \text{ ou } B(\{w,z\}) = B(\{v,u\}) = Y \cup \{\alpha\}.$$

   Les paires $\{w,z\}$ et $\{v,u\}$ ne sont pas séparées. On en déduit donc que soit u soit v est égal à α.

7. *Les paires $\{x,t\}$ et $\{z,w\}$ ne sont pas séparées.*

   D'après le résultat précédent, t est différent de α. On a :

   $$B(\{x,t\}) \cap Y = B(\{z,w\}) \cap Y = Y.$$

   Rappelons que les sommets x, y, et w n'ont aucun voisin hors de Y (lemme C.2 et résultat 1). Pour séparer les paires $\{x,t\}$ et $\{z,w\}$, il faut que t ou z ait un voisin hors de Y qui les sépare.

   Supposons d'abord que t ait un voisin s hors de Y qui sépare les paires $\{x,t\}$ et $\{z,w\}$; d'après la proposition C.1, t n'étant pas sommet d'articulation, il y a une (H,s,Y,t)-chaîne, C, qui ne peut arriver ni en v, ni en w, car on aurait un $\mathcal{C}_{\geq 7}$; elle ne peut arriver non plus ni en x, ni en y, car ces sommets n'ont pas de voisins n'appartenant pas à Y. Supposons que C arrive en u, voir figure C.IX(a) ; cela signifie que C se réduit à la chaîne u, s, t (sinon, existence d'un $\mathcal{C}_{\geq 7}$), et, d'après les hypothèses sur L ou d'après le résultat 6, s est différent de α ; d'après le lemme C.2 appliqué à w et s, il faut que w ou s soit adjacent à z. Supposons qu'il s'agisse de w ; on a :

   $$B(\{t,y\}) = B(\{t,w\}) = \{y,z,t,u,v,w\} \cup B(t).$$

   Comme y et w n'ont aucun voisin hors de Y, seul x pourrait séparer $\{t,y\}$ et $\{t,w\}$, mais on sait déjà que les seuls voisins de x dans G sont z et u. Les paires de sommets $\{t,y\}$ et $\{t,w\}$ ne sont donc pas séparées. Le sommet w ne peut donc pas être adjacent à z ; on montrerait de même que si s était adjacent à z, les paires $\{t,y\}$ et $\{t,s\}$ ne seraient pas séparées. La chaîne C ne peut donc pas arriver en u, et ne peut alors plus arriver qu'en z, et elle est de longueur 1 : voir figure C.IX(b), où s et z sont voisins. Ceci contredit toutefois le choix de s, qui devait séparer $\{x,t\}$ et $\{z,w\}$.

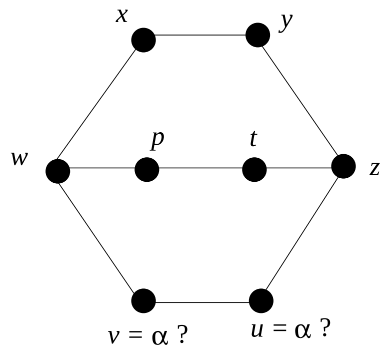


Figure C.X – *Le graphe K du lemme C.4.*

Supposons maintenant que z ait un voisin s hors de Y qui sépare les paires $\{x, t\}$ et $\{z, w\}$; d'après la proposition C.1, z n'étant pas sommet d'articulation, il y a une (H, s, Y, z)-chaîne C, qui ne peut arriver ni en v, ni en x, ni en y, sinon on aurait un $\mathcal{C}_{\geq 7}$; d'après le résultat 1, elle ne peut pas non plus arriver en w. Supposons que la chaîne C arrive en u ; elle serait alors de longueur 1 et, le sommet s n'étant pas le sommet d'articulation, cela contredirait le résultat 3. La chaîne C arrive donc en t et elle est de longueur 1 ; s et t sont voisins, ce qui contredit à nouveau le choix de s.

On ne peut pas séparer les paires $\{x, t\}$ et $\{z, w\}$.

L'hypothèse que H ne possède pas de $\mathcal{C}_{\geq 7}$ est contredite, et la démonstration du lemme C.3 est ainsi achevée. □

Lemme C.4. *On considère le graphe K représenté par la figure C.X et on suppose que, si α existe, il est en u ou v. Si le graphe H possède le graphe K comme sous-graphe partiel, alors H possède un cycle de longueur au moins 7 comme sous-graphe partiel.*

Preuve. On note Y l'ensemble des 8 sommets de K. On suppose que le graphe H possède le graphe K comme sous-graphe partiel et que, si α existe, il est en u ou v.

Le graphe G étant sans $(1, \leq 2)$-jumeaux, les ensembles $\{x, t\}$ et $\{y, p\}$ sont séparés. Par symétrie entre $\{x, y\}$ et $\{p, t\}$, puis par symétrie entre x et y, il suffit de supposer que le sommet x possède un voisin qui n'appartient pas à $B(\{y, p\})$. Or $B(\{y, p\}) \supseteq \{x, y, z, p, t, w\}$. On a alors les possibilités suivantes :

- x est voisin de $s \in X \setminus Y$, $s \neq \alpha$. Comme $x \neq \alpha$, il existe une (H, s, Y, x)-chaîne C. Si C arrive en w, y, p, t, v, ou u, alors on a un $\mathcal{C}_{\geq 7}$; et si C arrive en z, alors soit on a directement un $\mathcal{C}_{\geq 7}$, soit C est de longueur 1, ce qui signifie que les arêtes $\{x, s\}$ et $\{s, z\}$ existent, avec $y \neq \alpha, s \neq \alpha$, et on peut appliquer le lemme C.3.
- $\{x, v\}$ est une arête ou $\{x, u\}$ est une arête. Dans les deux cas, il existe un $\mathcal{C}_{\geq 7}$.

Dans tous les cas ci-dessus, on obtient un $\mathcal{C}_{\geq 7}$, et le lemme C.4 est ainsi démontré. □

Lemme C.5. *On considère le graphe K' représenté sur la figure C.XI et on suppose que, si α existe, il est en u ou v. Si le graphe H possède le graphe K' comme sous-graphe partiel, alors H possède un cycle de longueur au moins 7 comme sous-graphe partiel.*

Preuve. On note Y l'ensemble des 7 sommets de K' ; on suppose que le graphe H admet le graphe K' comme sous-graphe partiel et que, si α existe, il est en u ou v. Le graphe G étant sans $(1, \leq 2)$-jumeaux, les ensembles $\{p, x\}$ et $\{p, y\}$, dont les boules contiennent toutes deux les sommets x, y, z, w, et p, sont séparés ; on peut, sans perte de généralité,

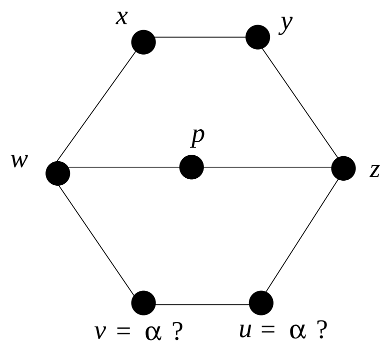


Figure C.XI – *Le graphe K' du lemme C.5.*

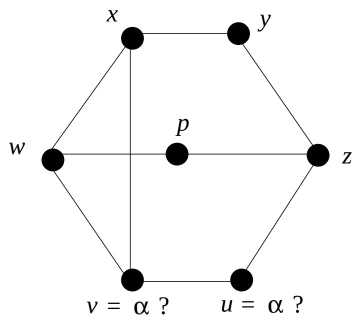


Figure C.XII – *Illustration pour la preuve du lemme C.5, avec l'arête $\{x, v\}$.*

supposer que le sommet x admet un voisin qui n'est pas dans $B(\{p, y\})$. On a alors les possibilités suivantes :

- (a) x est voisin de $s \in X \setminus Y$, $s \neq \alpha$. Comme $x \neq \alpha$, il existe une (H, s, Y, x)-chaîne C. Si C arrive en w, y, p, v, ou u, alors on a un $\mathcal{C}_{\geq 7}$; et si C arrive en z, alors soit on a directement un $\mathcal{C}_{\geq 7}$, soit C est de longueur 1 et on peut appliquer le lemme C.3, cf. preuve du lemme C.4.
- (b) $\{x, u\}$ est une arête ; alors il existe un $\mathcal{C}_{\geq 7}$.
- (c) $\{x, v\}$ est une arête, voir figure C.XII ; les ensembles $\{z, x\}$ et $\{z, w\}$, dont les boules contiennent tous les sommets de Y, étant séparés, le sommet w ou le sommet x doit avoir un voisin n'appartenant pas à Y. S'il s'agit de x, l'étude faite ci-dessus en (a) s'applique à nouveau. Examinons donc maintenant le sommet w, un voisin $s \in X \setminus Y$ de w qui ne soit voisin ni de x ni de z, et une (H, s, Y, w)-chaîne, C. Si C donne naissance à une chaîne de longueur 3 entre w et z dont seules les extrémités, w et z, sont dans Y, on applique le lemme C.4 ; dans tous les autres cas, on observe directement que H contient un $\mathcal{C}_{\geq 7}$.

Tous les cas conduisent à l'existence d'un cycle de longueur au moins 7. Le lemme C.5 est ainsi démontré. □

On peut maintenant montrer :

Lemme C.6. *Le plus long cycle de H n'est pas de longueur 6.*

Preuve. On procède par l'absurde en supposant que le plus long cycle de H est de longueur 6 ; si H possède un $\mathcal{C}_6$ contenant le sommet α, on choisit un tel cycle, sinon on choisit un $\mathcal{C}_6$ quelconque. On note a, b, c, d, e, et f les sommets de ce cycle, et on pose $Y = \{a, b, c, d, e, f\}$. Si le cycle contient le sommet α, on suppose que le sommet α est en f (voir figure C.XIII). Les lemmes C.3, C.4, et C.5 ainsi que l'inexistence d'un $\mathcal{C}_{\geq 7}$ montrent

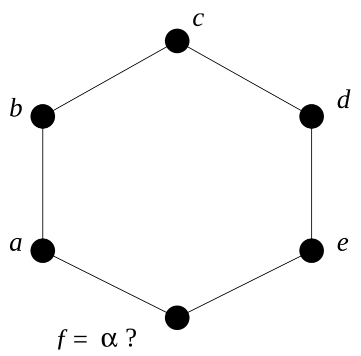


Figure C.XIII – *Le cycle de longueur 6 considéré pour le lemme C.6.*

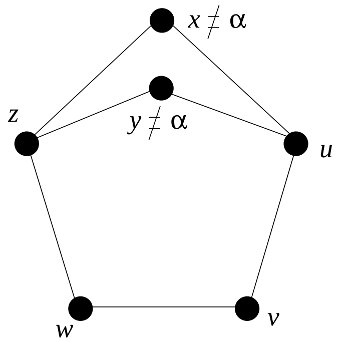


Figure C.XIV – *Le graphe M du lemme C.7.*

qu'il n'y a pas d'autre chaîne de longueur au moins 2 ayant ses extrémités dans Y et son (ses) autre(s) sommet(s) hors de Y que :

– une éventuelle chaîne de longueur 2 entre les sommets a et e ;
– une éventuelle chaîne de longueur 2 ou 3 entre les sommets c et f.

En effet, si une chaîne relie 2 sommets consécutifs du cycle, elle donne un $\mathcal{C}_{\geq 7}$; si elle relie 2 sommets à distance 2, autres que a et e, ou bien on a un $\mathcal{C}_{\geq 7}$, ou bien le lemme C.3 s'applique ; si elle relie 2 sommets opposés autres que c et f, soit elle donne un $\mathcal{C}_{\geq 7}$, soit le lemme C.4 ou C.5 s'applique ; enfin si elle est de longueur au moins 4 entre c et f, alors H contient un $\mathcal{C}_{\geq 7}$.

Les boules des ensembles $\{a, d\}$ et $\{b, e\}$ contiennent Y ; ces ensembles ne sont pas séparés, car on vient de voir que b et d n'ont pas de voisin hors de Y, et que a et e soit n'ont pas de voisin hors de Y, soit ont un seul voisin hors de Y, voisin qui leur est commun. □

C.4 Le plus long cycle de H n'est pas de longueur 5

Lemme C.7. *Si le graphe H possède le graphe M représenté par la figure C.XIV comme sous-graphe partiel avec $x \neq \alpha$ et $y \neq \alpha$, alors H possède un cycle de longueur au moins 6 comme sous-graphe partiel.*

Preuve. Supposons que H admette le graphe M, avec $x \neq \alpha$ et $y \neq \alpha$, comme sous-graphe partiel. Les ensembles $\{z, x\}$ et $\{z, y\}$ étant séparés, il faut que le sommet x ou le sommet y ait un voisin s assurant cette séparation. Supposons, sans perte de généralité, qu'il s'agisse du sommet x. S'il y a une arête entre x et v ou w, on a un $\mathcal{C}_{\geq 6}$. Sinon, le sommet x a un

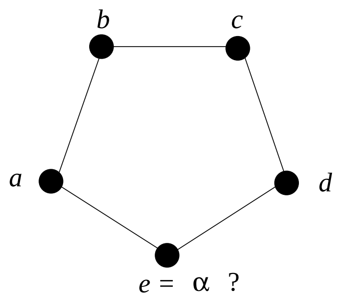


Figure C.XV – *Le cycle de longueur 5 considéré pour le lemme C.8.*

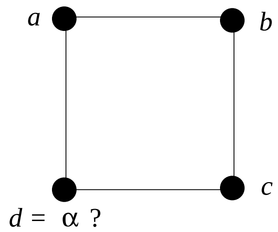


Figure C.XVI – *Le cycle de longueur 4 considéré pour le lemme C.9.*

voisin s n'appartenant pas à M ; le sommet x n'étant pas sommet d'articulation, il existe une (H, s, M, x)-chaîne qui dans tous les cas va engendrer un $\mathcal{C}_{\geq 6}$. □

Lemme C.8. *Le plus long cycle de H n'est pas de longueur 5.*

Preuve. On suppose que le plus long cycle de H est de longueur 5 et donc qu'il n'y a pas de $\mathcal{C}_{\geq 6}$. S'il existe un $\mathcal{C}_5$ contenant l'éventuel sommet d'articulation α, on choisit ce cycle et sinon on choisit un $\mathcal{C}_5$ quelconque ; on nomme a, b, c, d, e les sommets du cycle, et, si le cycle contient α, on suppose que $\alpha = e$ (voir figure C.XV).

Comme précédemment, l'inexistence d'un $\mathcal{C}_{\geq 6}$ et le lemme C.7 montrent que la seule chaîne possible de longueur au moins 2 dont les extrémités sont dans le cycle et les autres sommets n'appartiennent pas au cycle est une chaîne de longueur 2 entre a et d. Cela ne permet pas de séparer les ensembles $\{a, c\}$ et $\{b, d\}$, ce qui achève la preuve du lemme C.8, grâce au fait que a, c, b, d ne sont pas point d'articulation. □

C.5 Le plus long cycle n'est pas de longueur 4 ou 3

Lemme C.9. *Le plus long cycle de H n'est pas de longueur 4.*

Preuve. Supposons que le plus long cycle de H soit de longueur 4. On choisit un tel cycle, on nomme a, b, c, d ses sommets et on suppose sans perte de généralité que le sommet d'articulation n'est ni en a, ni en b, ni en c (voir figure C.XVI).

Les ensembles $\{b, a\}$ et $\{b, c\}$ étant séparés, il existe une chaîne de longueur au moins 2 dont une extrémité est a ou c, l'autre extrémité, distincte de la première, est sur le cycle et les autres sommets n'appartiennent pas au cycle ; la seule possibilité, pour ne pas avoir de $\mathcal{C}_{\geq 5}$, est une chaîne a, s, c où s n'appartient pas au cycle, mais alors le sommet s ne sépare pas les ensembles $\{b, a\}$ et $\{b, c\}$, ce qui achève la preuve du lemme C.9. □

Lemme C.10. *Le plus long cycle de H n'est pas de longueur 3.*

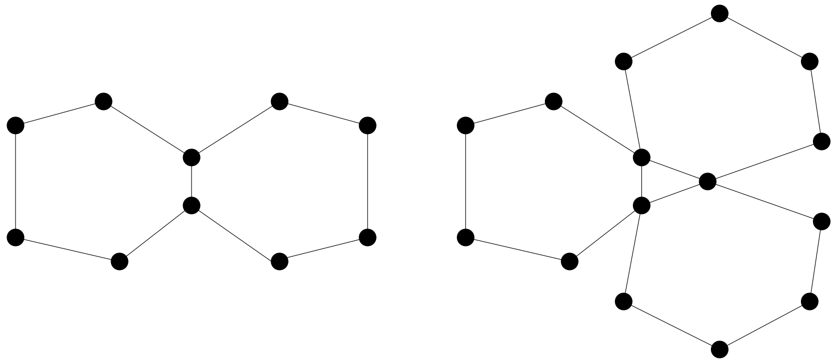


Figure C.XVII – *Deux graphes sans* $(1, \leq 2)$*-jumeaux et sans* $\mathcal{C}_{\geq 7}$ *sans corde.*

Preuve. Supposons que le plus long cycle de H soit de longueur 3 ; on choisit un $\mathcal{C}_3$ quelconque, on nomme a, b, c ses sommets et on suppose sans perte de généralité que l'éventuel sommet d'articulation n'est ni en a, ni en b. Il n'y a alors pas moyen de séparer les ensembles $\{c, a\}$ et $\{c, b\}$ sans créer de $\mathcal{C}_{\geq 4}$, d'où une contradiction. □

C.6 Existence d'un cycle de longueur au moins 7

Theorem C.2. *Tout graphe non orienté connexe d'ordre au moins 2 sans* $(1, \leq 2)$*-jumeaux possède un cycle élémentaire de longueur au moins 7 comme sous-graphe partiel.*

Preuve. On a vu, peu avant la Section C.3, que le graphe H admet un cycle ; d'après les lemmes C.6,C.8–C.10, son plus long cycle ne peut pas être de longueur 6, 5, 4, ou 3 : le plus long cycle de H, et donc le plus long cycle de G, est de longueur au moins 7. □

C.7 Conclusion : remarques et questions ouvertes

On va mettre en parallèle certains résultats obtenus sur les graphes sans $(1, \leq 2)$-jumeaux et les graphes sans $(r, \leq 1)$-jumeaux, en laissant ouvertes ces questions dans le cas des graphes sans $(r, \leq \ell)$-jumeaux, $r \geq 1$, $\ell \geq 1$.

On vérifie facilement qu'un graphe réduit à un cycle de longueur 7 (ou plus) est sans $(1, \leq 2)$-jumeaux, et qu'un graphe constitué d'un $\mathcal{C}_7$ et d'une ou plusieurs cordes possède des $(1, \leq 2)$-jumeaux. Ainsi,

- tout graphe connexe d'ordre au moins 2 sans $(1, \leq 2)$-jumeaux est d'ordre au moins 7 et le seul graphe connexe d'ordre 7 sans $(1, \leq 2)$-jumeaux est le graphe réduit à $\mathcal{C}_7$.

De manière similaire, on sait que

- tout graphe connexe d'ordre au moins 2 sans $(r, \leq 1)$-jumeaux est d'ordre au moins $2r + 1$ et le seul graphe connexe d'ordre $2r + 1$ sans $(r, \leq 1)$-jumeaux est la chaîne à $2r + 1$ sommets [35],[74]],[3].

Considérons maintenant les deux graphes de la figure C.XVII. Ces graphes sont sans $(1, \leq 2)$-jumeaux et ne possèdent pas de $\mathcal{C}_{\geq 7}$ sans corde. Ainsi,

- on ne peut pas ajouter au théorème C.2 la propriété d'avoir un $\mathcal{C}_{\geq 7}$ comme sous-graphe *induit*.

Cela diffère du cas des $(r, \leq 1)$-jumeaux, dans lequel il a été montré que

- tout graphe connexe d'ordre au moins 2 sans $(r, \leq 1)$-jumeaux contient la chaîne à $2r + 1$ sommets comme sous-graphe *induit* [5].

Finalement, on peut observer que le plus petit cycle possible, $\mathcal{C}_3$, peut apparaître dans un graphe sans $(1, \leq 2)$-jumeaux, ainsi qu'en atteste par exemple le deuxième graphe de la figure C.XVII.

Annexe D

Complexity Results for Identifying Codes in Planar Graphs

David Auger [1], Irène Charon [1],
Olivier Hudry [1], Antoine Lobstein [2]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Abstract

Let G be a simple, undirected, connected graph with vertex set $V(G)$ and $\mathscr{C} \subseteq V(G)$ be a set of vertices whose elements are called *codewords*. For $v \in V(G)$ and $r \geq 1$, let us denote by $I_r^{\mathscr{C}}(v)$ the set of codewords $c \in \mathscr{C}$ such that $d(v,c) \leq r$, where the distance $d(v,c)$ is defined as the length of a shortest path between v and c. More generally, for $A \subseteq V(G)$, we define $I_r^{\mathscr{C}}(A) = \cup_{v \in A} I_r^{\mathscr{C}}(v)$, which is the set of codewords whose minimum distance to an element of A is at most r. If r and l are positive integers, $\mathscr{C}$ is said to be an $(r, \leq l)$-identifying code if one has $I_r^{\mathscr{C}}(A) \neq I_r^{\mathscr{C}}(A')$ whenever A and A' are distinct subsets of $V(G)$ with at most l elements. We consider the problem of finding the minimum size of an $(r, \leq l)$-identifying code in a given graph. It is already known that this problem is NP-hard in the class of all graphs when $l = 1$ and $r \geq 1$. We show that it is also NP-hard in the class of planar graphs with maximum degree at most three for all (r, l) with $r \geq 1$ and $l \in \{1, 2\}$. This shows, in particular, that the problem of computing the minimum size of an $(r, \leq 2)$-identifying code in a given graph is NP-hard.

Keywords : graph theory ; identifying codes ; planar graphs ; complexity ; NP-completeness ; NP-hardness.
2000 *Mathematics Subject Classification :* 68Q17, 05C99, 94B65.

D.1 Notation and definitions

By *graph* we mean an undirected graph without loops nor multiple edges. If G is a graph, we denote respectively by $V(G)$ and $E(G)$ the sets of vertices and edges of G. An edge $\{x, y\} \in E(G)$ with $x, y \in V(G)$ will be simply denoted by xy. We refer to [43] for

1. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France
2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

basic notions such as adjacent vertices, paths, cycles or the neighbourhood of a vertex. Let us recall the distinction between a *subgraph* and an *induced subgraph* : a subgraph of G is a graph H with $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$, whereas if $W \subseteq V(G)$, the subgraph *induced by* G *on* W is the graph $G[W]$ whose vertex set is W, and whose edges are all the edges $xy \in E(G)$ with x and y in W.

From now on, we consider only connected graphs ; for $v \in V$ and $r \in \mathbb{N}$, the *ball* of radius r centered at v is the set

$$B(v, r) = \{w \in V(G) : d(v, w) \leq r\}$$

where $d(v, w)$ denotes the number of edges in any shortest path between v and w, i.e. the *length* of any shortest path between v and w. Whenever $d(v, w) \leq r$, we say that v and w *r-cover* each other (or simply *cover* if there is no ambiguity). A vertex $c \in V(G)$ is said to *r-separate* (or simply *separate*) vertices x and y if c r-covers one of them and does not r-cover the other.

In this paper, what we call a *code* is simply a set of vertices $\mathscr{C} \subseteq V(G)$, and we refer to its elements as *codewords*. A code $\mathscr{C}$ is said to be *r-covering* (or *r-distance-dominating*) if every vertex $v \in V(G)$ is r-covered by at least one codeword $c \in \mathscr{C}$. A code $\mathscr{C}$ is said to be *r-separating* if for every pair of distinct vertices $x \neq y$ of G there exists a codeword $c \in \mathscr{C}$ which r-separates x and y.

An *r-identifying code* is a code which is both r-covering and r-separating. Equivalently, $\mathscr{C} \subseteq V(G)$ is an r-identifying code if all the sets

$$I_r^{\mathscr{C}}(v) = B(v, r) \cap \mathscr{C}$$

for $v \in V(G)$ are non-empty and different.

More generally, if $\mathscr{C}$ is a code and A is a subset of $V(G)$ we denote by $I_r^{\mathscr{C}}(A)$ the set of codewords which r-cover at least one element of A, i.e.

$$I_r^{\mathscr{C}}(A) = \bigcup_{v \in A} B(v, r) \cap \mathscr{C}.$$

If r and l are positive integers, an *$(r, \leq l)$-identifying code* is a code $\mathscr{C}$ such that the sets $I_r^{\mathscr{C}}(A)$ for all $A \subseteq V(G)$ with $|A| \leq l$ are different. Note that, in this case, as $I_r^{\mathscr{C}}(\emptyset) = \emptyset$, then $I_r^{\mathscr{C}}(A) \neq \emptyset$ for $1 \leq |A| \leq l$; thus an $(r, \leq 1)$-identifying code is simply an r-identifying code.

We recall that the *degree* of a vertex $v \in V(G)$ is the number $\delta(v)$ of vertices $w \in V(G)$ such that $vw \in E(G)$. The maximum degree of G is defined as

$$\Delta(G) = \max_{v \in V(G)} \delta(v).$$

Finally, a graph is *planar* if it can be drawn in the plane in such a way that its edges do not cross. For precise definitions and additional background about graphs, we refer once again to [43]. The reader will also need basic knowledge in algorithmic complexity such as polynomial reduction and NP-completeness ; for these notions we refer to [50].

D.2 Introduction and main results

The problem of finding an identifying code of minimum size in a graph has been introduced and studied in [65] ; the original motivation was fault detection in processor systems. It was shown in [34] that the computation of the minimum size of an r-identifying code in a given graph is NP-hard for any $r \geq 1$; furthermore it was proved in [52] that this problem is APX-hard for $r = 1$. In particular, this implies if $P \neq NP$ that there exists a constant $c > 1$ such that no polynomial algorithm gives an efficiency ratio better than c. Indeed there is, for all $r \geq 1$, a polynomial approximation algorithm which computes an r-identifying code with efficiency ratio $O(\log |V(G)|)$, but sublogarithmic ratios are intractable (see [88]). For a nearly comprehensive bibliography about identifying codes, see [73].

In this paper we prove NP-hardness results for the restriction of this same problem to the class of planar graphs with maximum degree at most three ; this class is quite restrictive, as connected graphs with maximum degree at most two are paths and cycles where the size of a minimum r-identifying code is known exactly in most cases (see [17], [55], [82] and [89]). We also study the problem of finding the minimum size of an $(r, \leq 2)$-identifying code and prove its NP-hardness in the class of planar graphs with maximum degree at most three for all $r \geq 1$, which of course implies its NP-hardness in the class of all graphs. These codes have been investigated in [70], [53] and [69], but to our knowledge at this day no complexity result is known about them : our results show, in particular, that the problem of computing the minimum size of an $(r, \leq 2)$-identifying code in a given graph is NP-hard.

Let us denote by Π_3 the class of planar graphs with maximum degree at most three, and let r and l be positive integers. The problem that we study is precisely the following one :

<u>MIN $(r, \leq l)$-ID-CODE IN Π_3</u>
- INSTANCE : a graph $G \in \Pi_3$ and an integer k ;
- QUESTION : is there an $(r, \leq l)$-identifying code $\mathscr{C}$ of G with $|\mathscr{C}| \leq k$?

Our results can be summarized in the following theorem :

Theorem D.1. *The problem* MIN $(r, \leq l)$-ID-CODE IN Π_3 *is NP-complete for* $l \in \{1, 2\}$ *and all* $r \geq 1$*.*

D.3 Proofs of the complexity results

D.3.1 The vertex cover problem

Let G be a graph. An edge $e = xy \in E(G)$ is said to be *covered* by a vertex $v \in V(G)$ if v and e are incident, i.e. if $v = x$ or $v = y$. A *vertex cover* in G is a code $\mathscr{C} \subseteq V(G)$ such that every edge of G is covered by a at least one codeword $c \in \mathscr{C}$. Equivalently, $\mathscr{C}$ is a vertex cover if

$$\forall e = xy \in E(G), \quad x \in \mathscr{C} \text{ or } y \in \mathscr{C}.$$

It is well known that the problem of finding the minimum cardinality of a vertex cover in a given graph is NP-hard ([64]) ; furthermore, it was proved in [49] that this problem remains NP-hard in the class of planar graphs whose maximum degree is at most three. More precisely, the following problem is NP-complete :

<u>MIN VERTEX COVER IN Π_3</u>
- INSTANCE : a graph $G \in \Pi_3$ and an integer k ;
- QUESTION : is there a vertex cover $\mathscr{C}$ of G with $|\mathscr{C}| \leq k$?

If r and l are fixed and a code $\mathscr{C}$ is given in a graph G, as all distances between vertices of G can be computed in polynomial time, we can also compute all the sets $I_r^{\mathscr{C}}(A)$ for $A \subseteq V(G)$ with $|A| \leq l$ and compare them in polynomial time, and thus check that $\mathscr{C}$ is an $(r, \leq l)$-identifying code : therefore the problem Min $(r, \leq l)$-ID-code in Π_3 belongs to NP. We will complete the proof of Theorem D.1 by showing that Min Vertex Cover in Π_3 polynomialy reduces to the problem Min $(r, \leq l)$-ID-code in Π_3 for all $r \geq 1$ and $l \in \{1, 2\}$; we will consider four cases depending on the values of r and l : $r = 1$, $l = 1$ (Section D.3.2), $r = 2$, $l = 1$ (Section D.3.3), $r \geq 3$, $l = 1$ (Section D.3.4), and $r \geq 1$, $l = 2$ (Section D.3.5).

D.3.2 Reduction for $r = 1$ and $l = 1$

In this section, we set $r = 1$ and $l = 1$. Let us consider an instance of Min Vertex Cover in Π_3, i.e. a planar graph G with maximum degree at most three and an integer k. We construct a graph G' by replacing every edge $e = xy$ of G by the structure C_e counting 9 vertices and 11 edges specified by Fig. D.I. Note that we do not consider x and y as elements of the structure C_e ; thus C_e is not a graph (since it contains edges without containing their ends), but we will denote by $V(C_e)$ the set of its nine vertices as we would have done if it were one. See also Fig. D.II for an example of transformation.

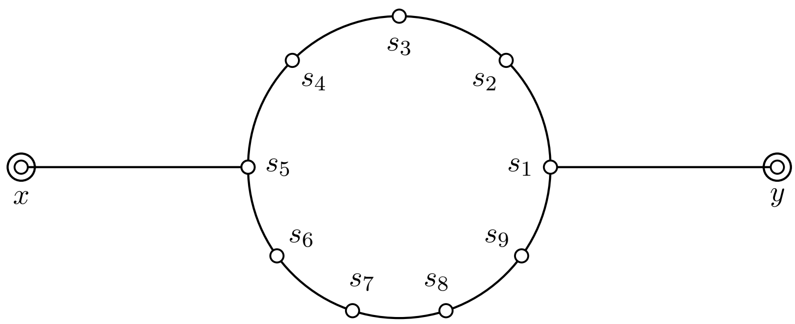


Figure D.I – *The structure C_e which replaces an edge $e = xy$ of G in G' .*

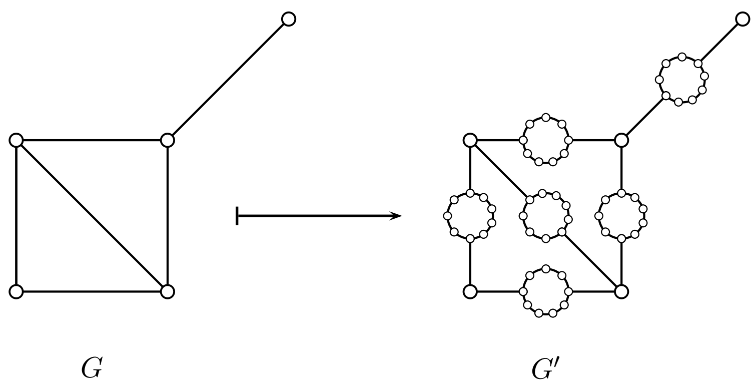


Figure D.II – *A graph G and the graph G' obtained through our transformation.*

If G has n vertices and m edges, then G' has $n + 9m$ vertices and $11m$ edges ; thus G'

can be constructed from G in polynomial time. It is easy to check that G' is also planar and has maximum degree three. The key-result for the reduction is the following :

Proposition D.2. *With notation above, G admits a vertex cover $\mathscr{C}$ with $|\mathscr{C}| \leq k$ if and only if G' admits a 1-identifying code $\mathscr{C}'$ with $|\mathscr{C}'| \leq k + 5m$.*

Proof. Let us fix our notation first : if G is the original graph and G' is the transformed one, we consider the vertex set $V(G)$ of G as a subset of $V(G')$. Therefore $V(G')$ can be partitioned in $|E(G)| + 1$ sets :

$$V(G') = V(G) \cup \bigcup_{e \in E(G)} V(C_e).$$

Let $\mathscr{C}$ be a vertex cover for G with $|\mathscr{C}| \leq k$; we construct an identifying code $\mathscr{C}'$ for G' by adding to $\mathscr{C}$ five vertices in each set $V(C_e)$ for all $e \in E(G)$; thus the code $\mathscr{C}'$ will have cardinality $|\mathscr{C}'| = |\mathscr{C}| + 5m \leq k + 5m$ as stated in the proposition. For each edge $e = xy \in E(G)$, the choice of the set $\mathscr{C}'_e$ of five vertices in $V(C_e)$ added to $\mathscr{C}$ will depend on the case $x \in \mathscr{C}$ or $y \in \mathscr{C}$; recall that as we have requested $\mathscr{C}$ to be a vertex cover of G, we have at least one of x and y in $\mathscr{C}$. As the structure is symmetric, without loss of generality let us assume that $x \in \mathscr{C}$. Then we define (see Fig. D.III)

$$\mathscr{C}'_e = \{s_1, s_2, s_5, s_6, s_9\}.$$

When this is done for all $e \in E(G)$, the code $\mathscr{C}'$ is defined as

$$\mathscr{C}' = \mathscr{C} \cup \bigcup_{e \in E(G)} \mathscr{C}'_e.$$

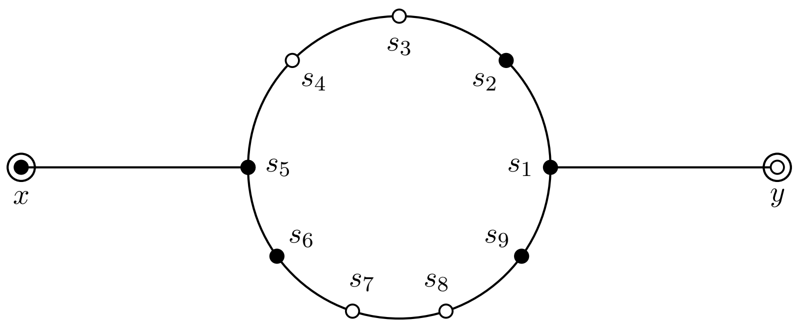


Figure D.III – *The code $\mathscr{C}'$ in C_e if $x \in \mathscr{C}$. Codewords are in black, non-codewords in white, with the exception of y which may be a codeword or not.*

It remains to prove that $\mathscr{C}'$ is a 1-identifying code ; since for each edge $e = xy \in E(G)$ the set of vertices covered by $\{s_1, s_2, s_5, s_6, s_9\}$ is precisely $V(C_e) \cup \{x, y\}$, it is easily seen that it suffices to check that for every edge $e = xy$ of G, all vertices in

$$V(C_e) \cup \{x, y\}$$

are covered and pairwise separated by $\mathscr{C}'$. We summarize in the following table which codewords cover the different vertices. Note that in C_e both x and y may belong to $\mathscr{C}$; this has no consequence since *containing* an identifying code is a sufficient condition for a set of vertices to *be* an identifying code.

vertices /codewords	x	s_1	s_2	s_5	s_6	s_9
x	•			•		
y		•				
s_1		•	•			•
s_2		•	•			
s_3			•			
s_4				•		
s_5	•			•	•	
s_6				•	•	
s_7					•	
s_8						•
s_9		•				•

Conversely, suppose that there exists a 1-identifying code $\mathscr{C}'$ for G' with $|\mathscr{C}'| \leq k+5m$. First we prove :
If neither x nor y belongs to $\mathscr{C}'$, then $|V(C_e) \cap \mathscr{C}'| \geq 6$. (D.1)
We just have to recall why a 1-identifying code on a cycle on 9 vertices must have at least 6 codewords. To simplify notation, let us consider that our vertices are integers modulo 9. If a is a vertex of the cycle, then $a+1$ and $a+2$ must be separated by a codeword, therefore a or $a+3$ must be codewords. The same argument proves that there must be a codeword in $\{a+3, a+6\}$, and another in $\{a+6, a\}$. So there must be at least two codewords in $\{a, a+3, a+6\}$ for each vertex a. Applied to $a \in \{1,2,3\}$, this argument shows that there are at least two codewords in each of the sets $\{1,4,7\}$, $\{2,5,8\}$ and $\{3,6,9\}$, therefore $|V(C_e) \cap \mathscr{C}'| \geq 6$.

Regardless of the fact that x or y belong to $\mathscr{C}'$, we have $|V(C_e) \cap \mathscr{C}'| \geq 5$. (D.2)
As in the previous case, there must be a codeword in $\{a, a+3\}$ for every a except in the cases when x or y can be used to separate $a+1$ from $a+2$, i.e. except for $a \in \{3,4,8,9\}$. So there must be at least one codeword in each of the pairs :

$$\{s_1, s_4\} \ \ \{s_2, s_5\} \ \ \{s_5, s_8\} \ \ \{s_6, s_9\} \text{ and } \{s_7, s_1\}.$$

We consider three cases :
- if neither s_1 nor s_5 are codewords, then s_4, s_2, s_8 and s_7 must be, as well as one of the pair $\{s_6, s_9\}$, so we need five codewords at least in $V(C_e)$;
- if s_1 and s_5 are both codewords : as at least one vertex in the pair $\{s_6, s_9\}$ is a codeword we can suppose by symmetry that $s_6 \in \mathscr{C}'$; but s_3 and s_8 are not covered by s_1, s_5, s_6, x nor y ; therefore we need at least two other codewords in $V(C_e)$;
- if s_1 is a codeword but s_5 is not (the other case being the same by symmetry), then s_2 and s_8 are codewords as well as one of the pair $\{s_6, s_9\}$. In both cases we still need to cover s_4 with a codeword in $V(C_e)$.

Let us define $\mathscr{C}$ as the trace of $\mathscr{C}'$ on $V(G)$, i.e.

$$\mathscr{C} = \mathscr{C}' \cap V(G).$$

Recall that $V(G)$ is a subset of $V(G')$ and so $\mathscr{C}$ is a code in G. We would like to use $\mathscr{C}$ in order to build a vertex cover of G. It may happen for an edge $e = xy \in E(G)$ that in G' neither x nor y belongs to $\mathscr{C}'$, and in this case the edge xy of G is not covered by the code $\mathscr{C}$. Let p be the number of edges $e \in E(G)$ which are not covered by $\mathscr{C}$. From (1) and (2) we have

$$|\mathscr{C}| \leq |\mathscr{C}'| - 6p - 5(m-p) = |\mathscr{C}'| - 5m - p \leq k - p.$$

As there are p uncovered edges, if we add to $\mathscr{C}$ one codeword on each edge uncovered by $\mathscr{C}$, we get a vertex cover of G with at most k codewords. □

D.3.3 Reduction for $r = 2$ and $l = 1$

For $r = 2$, we use a different strategy for the reduction. The basic idea is to replace every edge $e = xy \in E(G)$ by a chain on 4 vertices and 5 edges : see Fig. D.IV. Consider the central vertices a and b on the chain ; as a 2-identifying code has to separate a from b, we deduce that x or y must be a codeword. Hence the trace on G of a 2-identifying code of G' will be a vertex cover of G.

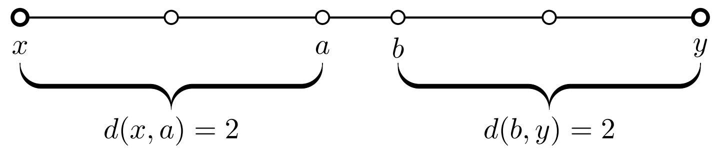


Figure D.IV – *A 2-identifying code will have to contain x or y to separate a from b.*

Clearly, this reduction is not entirely satisfactory because we cannot control the total number of codewords in the identifying code. In order to do so, we will add other devices around the chain.

Let $G \in \Pi_3$ and k be an integer. We construct from G a graph G' by replacing every edge $e = xy$ in $E(G)$ by a structure C_e counting 11 vertices and 14 edges : see Fig. D.V, where the chain $xv_5v_7v_8v_6y$ plays the role of the chain in Fig. D.IV. As in the previous case we do not consider x and y as elements of $V(C_e)$.

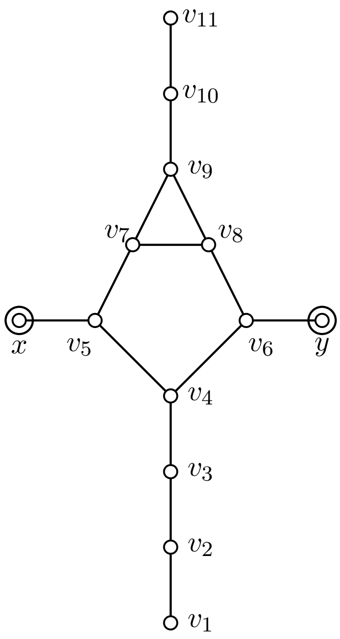


Figure D.V – *The structure C_e which replaces an edge $e = xy$ of G for $r = 2$.*

Clearly, G' can be constructed in polynomial time from G, and $G' \in \Pi_3$. Let $m = |E(G)|$. We will prove :

Proposition D.3. *G admits a vertex cover $\mathscr{C}$ with $|\mathscr{C}| \leq k$ if and only if G' admits a 2-identifying code $\mathscr{C}'$ with $|\mathscr{C}'| \leq k + 5m$.*

Proof. Once again, $V(G)$ is a subset of $V(G')$; so if we consider a vertex cover $\mathscr{C}$ of G with $|\mathscr{C}| \leq k$, we can add codewords to $\mathscr{C}$ in order to construct a 2-identifying code for G'. We do this in the following way : if $e = xy \in E(G)$, then x or y is in $\mathscr{C}$. Suppose without loss of generality that $x \in \mathscr{C}$. We add to $\mathscr{C}$ the 5 codewords v_3, v_4, v_6, v_8 and v_9 on the structure C_e (see Fig. D.VI).

When this is done for all $e \in E(G)$, we get a code $\mathscr{C}'$ on G' with $|\mathscr{C}'| = |\mathscr{C}| + 5m$, hence $|\mathscr{C}'| \leq k + 5m$. To check that $\mathscr{C}'$ is a 2-identifying code on G', we just have to see that for every $e = xy \in E(G)$, the vertices x, y and the 11 vertices of C_e are covered by different subsets of $\{x, v_3, v_4, v_6, v_8, v_9\}$. This is clearly sufficient because for a given $e = xy \in E(G)$, a vertex belongs to $V(C_e) \cup \{x, y\}$ if and only if it is 2-covered by v_3, v_4 or v_9. It remains to check the following table :

vertices /codewords	x	v_3	v_4	v_6	v_8	v_9
x	•		•			
y			•	•	•	
v_1		•				
v_2		•	•			
v_3		•	•	•		
v_4	•	•	•	•	•	
v_5	•	•	•	•	•	•
v_6		•	•	•	•	•
v_7	•		•	•	•	•
v_8			•	•	•	•
v_9				•	•	•
v_{10}					•	•
v_{11}						•

Conversely, suppose that $\mathscr{C}'$ is a 2-identifying code of G' with $|\mathscr{C}'| \leq k + 5m$. Then if $e = xy \in E(G)$, consider the codewords on C_e :

– v_1 and v_2 must be separated by $\mathscr{C}'$ so $v_4 \in \mathscr{C}'$;
– v_1 must be covered by $\mathscr{C}'$ so $\mathscr{C}' \cap \{v_1, v_2, v_3\} \neq \emptyset$;
– v_2 and v_3 must be separated by $\mathscr{C}'$ so $\mathscr{C}' \cap \{v_5, v_6\} \neq \emptyset$;
– v_{10} and v_{11} must be separated by $\mathscr{C}'$ so $\mathscr{C}' \cap \{v_7, v_8\} \neq \emptyset$;
– v_{11} must be covered by $\mathscr{C}'$ so $\mathscr{C}' \cap \{v_9, v_{10}, v_{11}\} \neq \emptyset$.

These five facts show that in each structure C_e, we have $|\mathscr{C}' \cap V(C_e)| \geq 5$, therefore the trace of $\mathscr{C}'$ on $V(G)$

$$\mathscr{C} = \mathscr{C}' \cap V(G)$$

has cardinality at most

$$|\mathscr{C}| \leq |\mathscr{C}'| - 5m \leq k.$$

Moreover, in each struture C_e, v_7 and v_8, which play the same role as a and b in Fig. D.IV, must be separated by $\mathscr{C}'$, so x or y must belong to $\mathscr{C}'$, hence to $\mathscr{C}$. Thus $\mathscr{C}$ is a vertex cover of G'. □

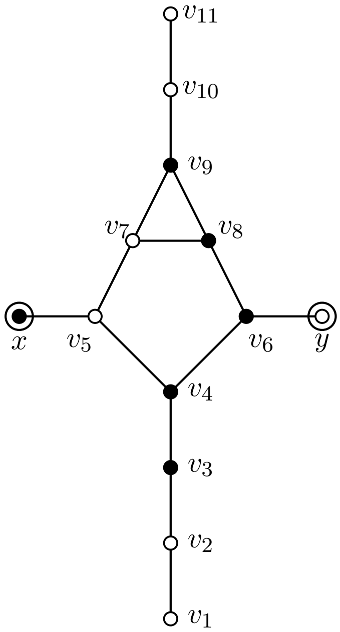


Figure D.VI – *The codewords in C_e for $r = 2$.*

D.3.4 Reduction for $r \geq 3$ and $l = 1$

For $r \geq 3$ our reduction uses the same idea as for $r = 2$, but the structure is slightly different. It is easier to present the proof for $r = 3$, but the general case is essentially the same.

Fix $r = 3$ and let $G \in \Pi_3$. We replace every edge $e = xy$ of $E(G)$ by the structure specified by Fig. D.VII. In particular, there are 5 cycles with attached paths on this structure that we call *suns*. A single sun is displayed on Fig. D.VIII. There are, in Fig. D.VII, 189 vertices, not counting x and y, and our transformation is polynomial; moreover, the new graph G' is clearly in Π_3.

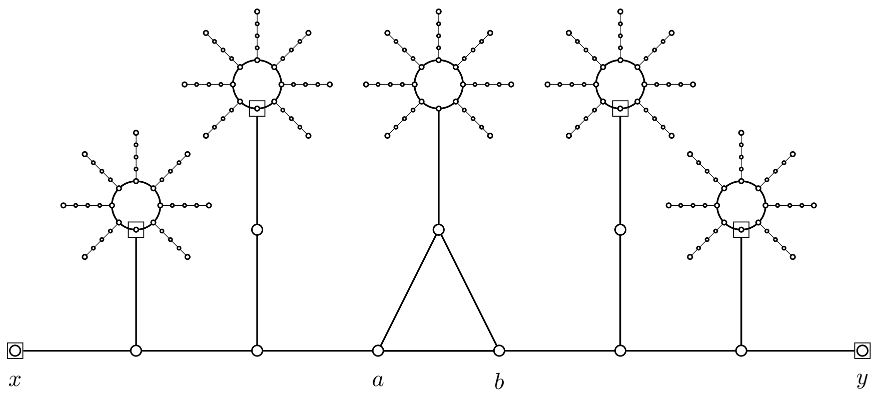


Figure D.VII – *The structure C_e which replaces an edge $e = xy$ of G for $r = 3$. Squared vertices belong to $B(a, 3) \Delta B(b, 3)$.*

Now we will prove the following proposition, where $m = |E(G)|$:

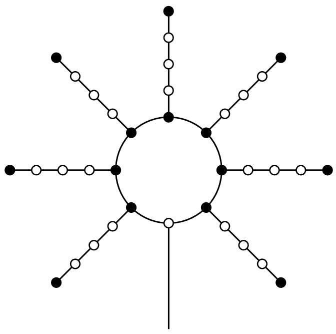


Figure D.VIII – *A sun and its codewords for $r = 3$.*

Proposition D.4. *G admits a vertex cover $\mathscr{C}$ with $|\mathscr{C}| \leq k$ if and only if G' admits a 3-identifying code $\mathscr{C}'$ with $|\mathscr{C}'| \leq k + 70m$.*

Proof. First suppose that we have a vertex cover $\mathscr{C}$ of G with $|\mathscr{C}| \leq k$. Then we construct a 3-identifying code $\mathscr{C}'$ on G' by adding 70 codewords to $\mathscr{C}$ on each structure C_e ; these codewords correspond to 14 codewords on each sun, as in Fig. D.VIII, thus $|\mathscr{C}'| \leq k + 70m$. We leave the readers to convince themselves that this defines a 3-identifying code on G' ; just note that as in the case $r = 2$, the central vertices a and b are separated because x or y belongs to $\mathscr{C}$ and therefore is a codeword of $\mathscr{C}'$. Once again, it will be sufficient to check that $\mathscr{C}'$ 3-separates and 3-covers vertices inside each structure C_e if we first observe that $\mathscr{C}'$ enables us, for any vertex $v \in V(G')$, just by looking at the set $I_r^{\mathscr{C}'}(v)$, to know if v belongs to $V(G)$ or to identify the structure C_e to which v belongs.

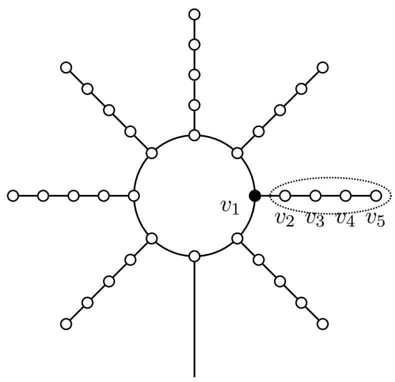


Figure D.IX – *Necessary codewords in a 3-identifying code in a sun.*

Conversely, if $\mathscr{C}'$ is a 3-identifying code on G' with $|\mathscr{C}'| \leq k + 70m$, then consider a single sun of a structure C_e (see Fig. D.IX). In order to separate vertices v_4 and v_5, v_1 must be a codeword. In order to cover v_5, we must have $\mathscr{C}' \cap \{v_2, v_3, v_4, v_5\} \neq \emptyset$. The same argument holds for every ray of the sun, therefore we need at least $2 \times 7 = 14$ codewords in it ; as there are five suns in C_e we need at least 70 codewords on C_e.

The code $\mathscr{C}'$ being 3-identifying on G', it necessarily separates the vertices a and b, i.e. there is at least one codeword in the symmetric difference $B(a, 3)\Delta B(b, 3)$. This set is composed of six vertices (see Fig. D.VII) : x, y and four vertices belonging to C_e. Note that none of these six vertices was counted above in the 70 codewords. Hence we have (recall that x and y do *not* belong to C_e) :

– if x or y belong to $\mathscr{C}'$ then $|\mathscr{C}' \cap V(C_e)| \geq 70$;
– if neither x nor y belongs to $\mathscr{C}'$ then $|\mathscr{C}' \cap V(C_e)| \geq 71$.

The conclusion goes exactly like in the case $r = 1$: let us define

$$\mathscr{C} = \mathscr{C}' \cap V(G)$$

which is a code on G and let p be the number of edges $e \in E(G)$ uncovered by $\mathscr{C}$. Then

$$|\mathscr{C}| = |\mathscr{C}' \cap V(G)| \leq |\mathscr{C}'| - 71p - 70(m-p) \leq k - p$$

because we supposed $|\mathscr{C}'| \leq k + 70m$. Thus if we add to $\mathscr{C}$ one codeword on each of the p edges of G uncovered by $\mathscr{C}$, we get a vertex cover of G, whose cardinal is at most k. □

For $r \geq 4$ our proof is exactly the same as in the case $r = 3$ but the structure has to be adapted. If we consider an edge $e = xy$ of G, we replace it by a path on $2r$ extra vertices and attach five suns to the path as on Fig. D.X. Each sun is now a cycle on $2r+2$ vertices with $2r+1$ rays of length $r+1$. With this structure replacing every edge of G, we obtain a graph G', satisfying the same properties as in the case $r = 3$. One can prove the following proposition exactly as it was done for $r = 3$.

Proposition D.5. *G admits a vertex cover $\mathscr{C}$ with $|\mathscr{C}| \leq k$ if and only if G' admits an r-identifying code $\mathscr{C}'$ with $|\mathscr{C}'| \leq k + 10(2r+1)m$.*

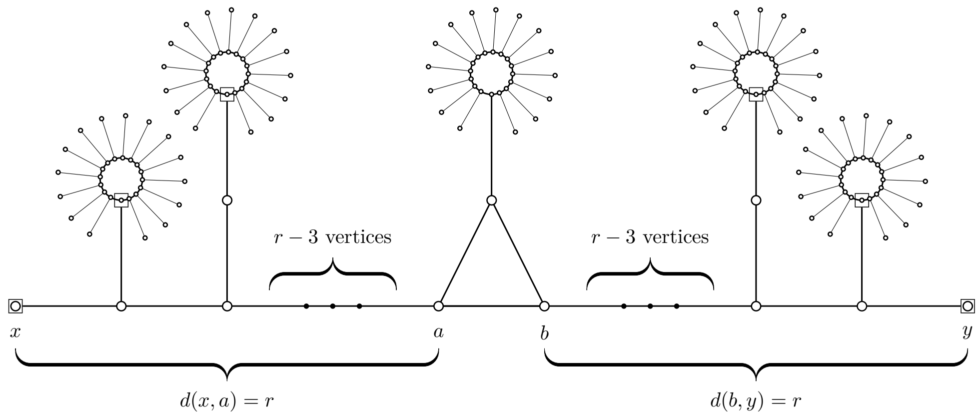


Figure D.X – *The structure C_e in the general case $r \geq 3$.*

D.3.5 Reduction for $r \geq 1$ and $l = 2$

For our transformation we need to define a restriction of the Min Vertex Cover in Π_3 problem. Let Π'_3 be the class of graphs G such that :

– G belongs to Π_3, i.e. G is planar and every vertex of G has degree at most three;
– no vertex of G has degree one;
– if v_1 and v_2 are distinct vertices of G with degree 2, there exist vertices v'_1 and v'_2, distinct from v_1 and v_2, such that v'_1 is adjacent to v_1 but not to v_2, and v'_2 is adjacent to v_2 but not to v_1.

We define a problem Min Vertex Cover in Π'_3 by analogy with the problem Min Vertex Cover in Π_3. We will prove the following result :

Lemma D.6. *The problem* Min Vertex Cover in Π'_3 *is NP-complete.*

Before proving this lemma let us begin with three preliminary results.

Lemma D.7. *Let G be a graph and $xy \in E(G)$ be an edge such that the degree of the vertex x is 1, and let $G' = G[V(G) \setminus \{x, y\}]$ be the graph obtained when we remove x, y and all their incident edges from G. Then the minimum cardinality of a vertex cover in G equals the minimum cardinality of a vertex cover in G' plus* 1 *(see Fig. D.XI).*

Proof. Suppose that $\mathscr{C}$ is a minimum vertex cover in G ; then exactly one of x and y belongs to $\mathscr{C}$, because if we had x and y in $\mathscr{C}$, then $\mathscr{C} \setminus \{x\}$ would also be a vertex cover of G. Hence $\mathscr{C}' = \mathscr{C} \setminus \{x, y\}$ is a vertex cover of G' and $|\mathscr{C}'| = |\mathscr{C}| - 1$. Conversely, if $\mathscr{C}'$ is a (minimum) vertex cover of G', then $\mathscr{C}' \cup \{y\}$ is a (minimum) vertex cover of G. $\square$

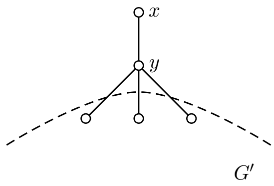


Figure D.XI – *Illustration of Lemma D.7.*

Lemma D.8. *Let G be a graph and x be a vertex of degree 2 whose neighbours y and z are adjacent, and let $G' = G[V(G) \setminus \{x, y, z\}]$ be the graph obtained when we remove x, y, z and all their incident edges from G. Then the minimum cardinality of a vertex cover in G equals the minimum cardinality of a vertex cover in G' plus* 2 *(see Fig. D.XII).*

Proof. Suppose that $\mathscr{C}$ is a minimum vertex cover in G ; then exactly two of the three vertices x, y and z belong to $\mathscr{C}$, because we need at least two of them to cover the edges xy, yz and zx, and if we had x, y and z in $\mathscr{C}$ then $\mathscr{C} \setminus \{x\}$ would still be a vertex cover of G. So $\mathscr{C}' = \mathscr{C} \setminus \{x, y, z\}$ is a vertex cover of G' and $|\mathscr{C}'| = |\mathscr{C}| - 2$. Conversely, if $\mathscr{C}'$ is a (minimum) vertex cover of G', then $\mathscr{C}' \cup \{y, z\}$ is a (minimum) vertex cover of G. $\square$

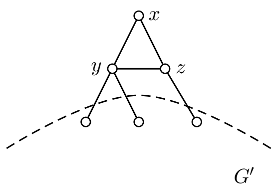


Figure D.XII – *Illustration of Lemma D.8.*

Note that the previous two lemmas could be generalized in the following way : if x is a simplicial vertex in G, i.e. a vertex x with degree $\delta \geq 1$ whose δ neighbours induce a clique in G, then the minimum cardinality of a vertex cover in G is equal to the mimimum cardinality of a vertex cover in $G[V \setminus B(x, 1)]$ plus δ. [1]

Lemma D.9. *Let G be a graph and x, y be two distinct vertices of G with degree* 2*, sharing the same neighbours z and z' (in particular x and y are non adjacent). Let $G' = G[V(G) \setminus \{x, y, z, z'\}]$ be the graph obtained when we remove x, y, z, z' and all their incident edges from G. Then the minimum cardinality of a vertex cover in G equals the minimum cardinality of a vertex cover in G' plus* 2 *(see Fig. D.XIII).*

1. We thank the referee for this remark.

Proof. If $\mathscr{C}$ is a minimum vertex cover in G, then exactly two of the four vertices x, y, z and z' belong to $\mathscr{C}$; the fact that $zz' \in E(G)$ or not does not matter. The conclusion follows as in the previous two lemmas. □

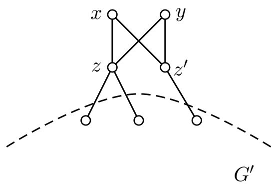


Figure D.XIII – *Illustration of Lemma D.9.*

Proof of Lemma D.6. Suppose that G belongs to Π_3, but not to Π'_3. Then either :

- there exists a vertex x of degree one in G, and we will apply Lemma D.7 ;
- there exist vertices $x \neq y$ of degree two such that every neighbour of x is either y or adjacent to y. Then :
  - if x and y are adjacent, let z be their common neighbour. Then we can apply Lemma D.8 ;
  - if x and y are non adjacent, let z and z' be their (common) neighbours ; we can apply Lemma D.9.

In the three cases, we obtain a 'reduced' graph G' and we can trivially compute the size of a minimum vertex cover in G if we know the size of a minimum vertex cover in G'. Clearly, the reduced graph G' is planar and has maximum degree at most three. If this graph does not belong to Π'_3, we can reduce it once again, and so on until we obtain an 'irreducible' graph G'', i.e. a graph which belongs to Π'_3. Thus we obtain by an algorithm which is obviously polynomial a graph $G'' \in \Pi'_3$ and an integer k such that the size of a minimum vertex cover in G is equal to the size of a minimum vertex cover in G'' plus k. This proves that the problem MIN VERTEX COVER IN Π'_3 is algorithmically harder than MIN VERTEX COVER IN Π_3. As the converse is obviously true and since MIN VERTEX COVER IN Π_3 is NP-complete, we have proved Lemma D.6. □

Let us now come back to $(r, \leq 2)$-identifying codes. Let $G \in \Pi'_3$; we replace every edge $e \in E(G)$ of G by a structure C_e which can be seen on Fig. D.XIV. The structure C_e is made of a cycle on $4r + 3$ vertices, attached to x and y by a path on $2r + 1$ vertices. As before, we do not consider that x and y belong to $V(C_e)$.

Clearly, if $G \in \Pi'_3$ then $G' \in \Pi_3$, and G' can be computed in polynomial time from G. The reduction of the problem MIN VERTEX COVER IN Π'_3 to the problem MIN $(r, \leq 2)$-ID-CODE IN Π_3 will be obtained if we prove :

Proposition D.10. *G admits a vertex cover $\mathscr{C}$ with $|\mathscr{C}| \leq k$ if and only if G' admits an $(r, \leq 2)$-identifying code $\mathscr{C}'$ with $|\mathscr{C}'| \leq k + (6r + 4)m$.*

In order to shorten the proof of Proposition D.10, we need two other technical lemmas :

Lemma D.11. *Let $r \geq 1$ and H be a graph consisting of a path on $3r + 1$ vertices*

$$p_{-r+1}\ p_{-r+2} \cdots p_{-1}\ p_0\ p_1\ p_2 \cdots p_{2r+1},$$

a cycle on $4r + 3$ vertices

$$c_1\ c_2 \cdots c_{4r+3}\ c_1$$

and the edge $p_{2r+1}c_1$ which connects the path to the cycle (see Fig. D.XV). Let W be the set of vertices

$$W = \{p_1, p_2, \cdots, p_{2r+1}\} \cup \{c_1, c_2, \cdots, c_{4r+3}\}$$

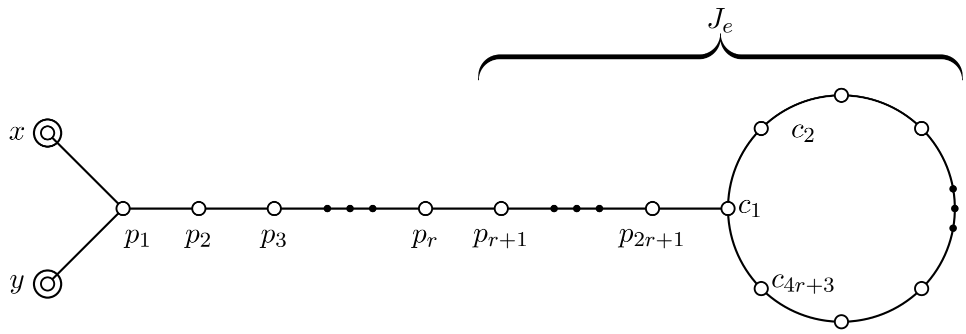


Figure D.XIV – *The structure C_e which replaces an edge $e = xy$ of G in the case $l = 2$.*

and let $\mathscr{C} = V(H)$ be the code consisting of all vertices in H. Then all the identifying sets $I_r^{\mathscr{C}}(A)$, for all $A \subset W$ with $|A| \leq 2$, are distinct.

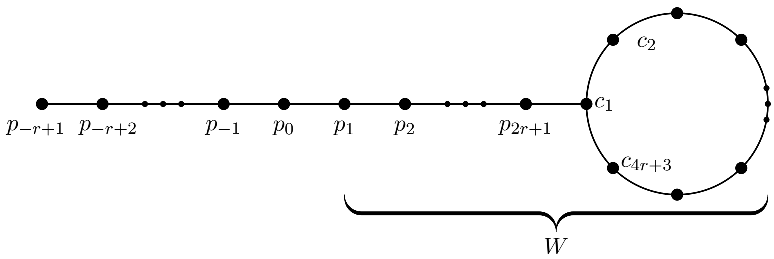


Figure D.XV – *The graph H in Lemma D.11.*

Proof. The lemma is proved if we show that $I_r^{\mathcal{C}}(A_1) \neq I_r^{\mathcal{C}}(A_2)$ in the 15 cases given by the Table below.

$A_1 =$	$\{p_{i_1}\}$	$\{c_{i_1}\}$	$\{p_{i_1}, p_{i_2}\}$ $(i_1 < i_2)$	$\{c_{i_1}, c_{i_2}\}$	$\{p_{i_1}, c_{i_2}\}$
$A_2 = \{p_{j_1}\}$	1	2	3	4	5
$A_2 = \{c_{j_1}\}$		6	7	8	9
$A_2 = \{p_{j_1}, p_{j_2}\}$ $(j_1 < j_2)$			10	11	12
$A_2 = \{c_{j_1}, c_{j_2}\}$				13	14
$A_2 = \{p_{j_1}, c_{j_2}\}$					15

The cases $2, 4, 7, 9, 11, 14$ are easy to check : in cases $2, 4, 11$, the codeword $p_{j_1 - r}$ r-covers p_{j_1} and neither c_{i_1} nor c_{i_2}, so $I_r^{\mathcal{C}}(A_1) \neq I_r^{\mathcal{C}}(A_2)$; similarly, in cases $7, 9, 14$, $p_{i_1 - r}$ covers p_{i_1}, and neither c_{j_1} nor c_{j_2}.

In case 1, assuming without loss of generality that $j_1 < i_1$, $p_{j_1 - r}$ covers p_{j_1}, not p_{i_1}.

In case 3, if $j_1 = i_1$, then either $p_{i_2 + r}$, if it exists, or a codeword of type c covers p_{i_2}, not p_{j_1} ; otherwise, set $\mu = \min\{i_1, j_1\}$ and consider $p_{\mu - r}$.

In cases $5, 12$, one of $c_{i_2 \pm r \mod 4r+3}$ covers c_{i_2}, and neither p_{j_1} nor p_{j_2}.

Cases $6, 8$ are also easy to handle.

In case 10, consider $p_{\min\{i_1,j_1\}-r}$ if $i_1 \neq j_1$; if $i_1 = j_1$, then $p_{\max\{i_2,j_2\}+r}$, if it exists, or a codeword of type c will do.

Case 15 is easy, both if $i_1 = j_1$ and if $i_1 \neq j_1$.

We are left with case 13, where, instead of showing that $I_r^{\mathcal{C}}(\{c_{i_1}, c_{i_2}\}) \neq I_r^{\mathcal{C}}(\{c_{j_1}, c_{j_2}\})$, we equivalently show that, given $I_r^{\mathcal{C}}(\{c_1, c_2\})$, we can uniquely recover c_1 and c_2.

It is quite straightforward to observe that, with computations carried modulo $4r+3$, the set $I_r^{\mathcal{C}}(\{c_1, c_2\})$ is of the form $\{c_\alpha, c_{\alpha+1}, \cdots, c_{\alpha+\beta}\}$, plus maybe vertices of type p, where $2r+1 \leq \beta \leq 4r+1$. Then $\{c_{\alpha+r}, c_{\alpha+\beta-r}\} = \{c_1, c_2\} \mod 4r+3$. ☐

Lemma D.12. *Let $r \geq 1$ and H be a graph consisting of a path on $2r+1$ vertices*

$$p_1\ p_2 \cdots p_{2r+1},$$

a cycle on $4r+3$ vertices

$$c_1\ c_2 \cdots c_{4r+3}\ c_1$$

and the edge $p_{2r+1}c_1$ which connects the path to the cycle (see Fig. D.XVI). Let $\mathscr{C}$ be the code on H defined by

$$\mathscr{C} = \{p_{r+1}, p_{r+2}, \cdots, p_{2r+1}\} \cup \{c_1, c_2, \cdots, c_{4r+3}\}.$$

Then $\mathscr{C}$ is an r-identifying code on H.

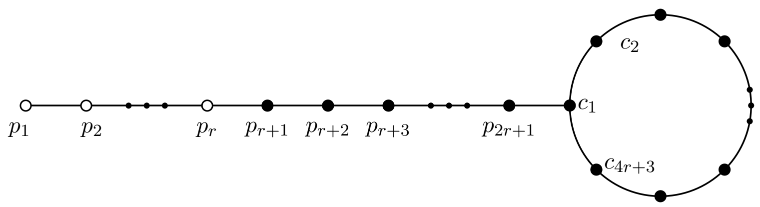


Figure D.XVI – *The graph H in Lemma D.12. Codewords are in black.*

Proof. The proof being much simpler than the proof of Lemma D.11, we leave it to the readers who will easily convince themselves that the sets $I_r^{\mathscr{C}}(v)$ are all non-empty and distinct for all $v \in V(H)$. ☐

Proof of Proposition D.10. Consider a vertex cover $\mathscr{C}$ of G with $|\mathscr{C}| \leq k$. As in the previous sections, $V(G)$ is a subset of $V(G')$, and so $\mathscr{C}$ is a subset of $V(G')$. We construct an $(r, \leq 2)$-identifying code $\mathscr{C}'$ of G' by adding to $\mathscr{C}$ all the vertices in $V(C_e)$ for each $e \in E(G)$ (cf. Fig. D.XIV). As each structure C_e counts $6r+4$ vertices, the code $\mathscr{C}'$ will have cardinality

$$|\mathscr{C}'| = |\mathscr{C}| + (6r+4)m \leq k + (6r+4)m.$$

Before checking that $\mathscr{C}'$ is an $(r, \leq 2)$-identifying code of G', let us recall that $V(G')$ is partitioned in the following way :

$$V(G') = V(G) \cup \bigcup_{e \in E(G)} V(C_e).$$

Consider a set of vertices $A \subseteq V(G')$ with $|A| \in \{1,2\}$; if we are given $I_r^{\mathscr{C}'}(A)$, we want to identify A. If $e \in E(G)$ let us define J_e as the following subset of $V(C_e)$ (see Fig. D.XIV) :

$$J_e = V(C_e) \setminus \{p_1, \cdots, p_r\}.$$

We will repeatedly use the following obvious fact :

$$I_r^{\mathscr{C}'}(A) \cap J_e \neq \emptyset \text{ if and only if } A \cap V(C_e) \neq \emptyset. \tag{D.3}$$

Consider three cases :

- *First case : there exist two distinct edges e_1, e_2 in $E(G)$ such that*

$$I_r^{\mathscr{C}'}(A) \cap J_{e_i} \neq \emptyset \textit{ for } i \in \{1,2\}.$$

By (D.3) we must have $|A| = 2$ and A consists of two vertices v_1 and v_2 respectively belonging to $V(C_{e_1})$ and $V(C_{e_2})$. Since $e_1 \neq e_2$, we have $I_r^{\mathscr{C}'}(v_2) \cap J_{e_1} = \emptyset$; if we denote by H be the graph induced by G' on $V(C_{e_1})$ (cf. Fig. D.XVI), Lemma D.12 implies that J_{e_1} is an r-identifying code on H and thus v_1 can be identified. By the same argument v_2, and thus A can be identified.

- *Second case : there is a single edge $e_1 \in E(G)$ such that $I_r^{\mathscr{C}'}(A) \cap J_{e_1} \neq \emptyset$.*

In this case there are three possibilities :

– either $|A| = 2$ and A consists of two vertices x and v_1, where $x \in V(G)$ and $v_1 \in V(C_{e_1})$;
– or $|A| = 2$ and A consists of two vertices $v_1, v_2 \in V(C_{e_1})$;
– or $|A| = 1$ and A consists of a single vertex $v_1 \in V(C_{e_1})$.

We can simply detect if a vertex $x \in V(G)$ belongs to A : such a vertex x is r-covered by the vertices p_r in each structure C_e such that e is incident with x in G. Suppose that $x \in A$. Since $G \in \Pi_3'$, x has a degree at least 2 in G, and so there must exist two vertices p_r and p_r' in two distinct structures C_e and $C_{e'}$ such that p_r and p_r' belong to $I_r^{\mathscr{C}'}(A)$. If this happens, we easily find x as the common endpoint of e and e'. Conversely, if $A \cap V(G) = \emptyset$, we know that $A \subset V(C_{e_1})$ and the vertices p_r in structures C_e with $e \neq e_1$ cannot belong to $I_r^{\mathscr{C}'}(A)$.

Thus we have proved that in the above three possibilities, we know if we are in the first one and then we can identify x. It will remain to identify the vertex $v_1 \in C_{e_1}$; as x does not r-cover the vertices of J_{e_1}, we know by Lemma D.12 that this can be done.

In the case $A \cap V(G) = \emptyset$, we must determine if $|A| = 1$ or $|A| = 2$, and then find A. Let $e_1 = xy$, with $x, y \in V(G)$; since $\mathscr{C}$ is a vertex cover of G we know that x or y belong to $\mathscr{C}$, and thus to $\mathscr{C}'$. Let us suppose without loss of generality that $x \in \mathscr{C}'$. The degree of x in G is at least two, therefore there must exist an edge $e_2 = xz$ incident with e_1 in x. Let us denote by $p_1', p_2', \cdots, p_r'$ the vertices on the path of the structure C_{e_2}. If we consider the union of the path $p_{r-1}' \cdots p_1'$ in C_{e_2}, with the vertex x and the set $V(C_{e_1})$ (see Fig. D.XVII), we find an (isometric) induced subgraph of G' where every vertex is a codeword. Since we know that $A \subset V(C_{e_1})$, Lemma D.11 can be applied, with x playing the role of p_0, and thus A can be identified.

- *Third case : for all $e \in E(G)$ we have $I_r^{\mathscr{C}'}(A) \cap J_e = \emptyset$.*

Then by (D.3), A consists of one or two vertices which belong to $V(G)$. Let us denote by F the set of edges $e = xy \in E(G)$ such that the vertex p_r of C_e belongs to $I_r^{\mathscr{C}'}(A)$. If $e = xy \in F$, with $x, y \in V(G)$, we know that $x \in A$ or $y \in A$, and the converse is also true : thus F is precisely the set of edges of G which are incident in G with the vertex, or the two vertices, in A. Can we find A if F is given ?

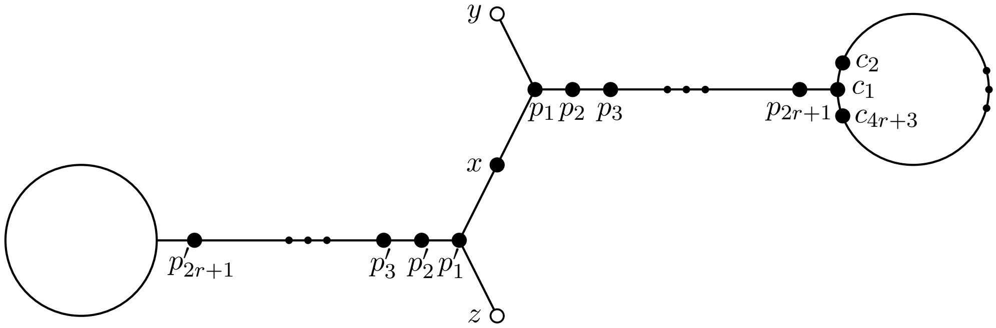


Figure D.XVII – *An induced subgraph where Lemma D.11 can be applied.*

Let us consider the graph $G[F]$ whose edge set is F and whose vertex set consists of the elements of $V(G)$ incident with at least one edge in F (this is not always an induced subgraph of G). We claim that we can identify A by looking at the degrees of its vertices in G and in $G[F]$.
Note that if $z \in A$, then z has the same degree in $G[F]$ as in G, and this degree is two or three because $G \in \Pi'_3$; and if $z \in V(G[F]) \setminus A$, then z, in $G[F]$, can have neighbours only in A, so its degree is one or two.
Thus if $|A| = 1$, every vertex of $G[F]$ has degree one in $G[F]$ with the exception of the vertex x such that $A = \{x\}$, whereas if $|A| = 2$ there must exist at least two vertices in $G[F]$ with degree at least two. Therefore by counting the number of vertices with degree at least two in $G[F]$, we can tell if $|A|$ is equal to 1 or 2, and identify A in the former case.

If $|A| = 2$, let $A = \{x, y\}$. Because $G \in \Pi'_3$, we have three cases :

(i) If there is no vertex of degree three in $G[F]$, we have $\delta(x) = \delta(y) = 2$; then there must exist vertices z and z', distinct from x and y, such that z is adjacent to x and not to y and z' is adjacent to y and not to x. Thus z and z' are vertices of $G[F]$ and have degree one in $G[F]$: this enables us to identify x and y as the only vertices adjacent to vertices of degree one in $G[F]$.

(ii) If there are two vertices with degree three in $G[F]$, we have $\delta(x) = \delta(y) = 3$: these are precisely the only two vertices with degree three in $G[F]$, and they can be identified as such.

(iii) We are left with the case when, without loss of generality, $\delta(x) = 3$ and $\delta(y) = 2$; then x is identified as the only vertex with degree three in $G[F]$.
If at least one of the neighbours of y has degree one in $G[F]$, then, as in case (i), y can be identified by the fact that it is adjacent to a vertex of degree one in $G[F]$. So now we assume that the two neighbours of y, z_1 and z_2, have degree two or three in $G[F]$; note that if z_i has degree two in $G[F]$ then it is adjacent to x and y in $G[F]$, and if it has degree three in $G[F]$ then $z_i = x$.
If x and y are non adjacent, then z_1, z_2 are adjacent to x, y, and y can be identified as the only vertex in $G[F]$ not adjacent to x (which has already been identified). If x and y are adjacent, then we have the following edges in $G[F]$: $yx = yz_1$, xz_2, yz_2, xz_3. Now how to know whether it is y or z_2 which belongs to A? They cannot both have degree two in G, because the characterization of Π'_3 states that they should have distinct neighbours. So z_2 has degree three in G, and y is identified by the fact that it has degree two in G and in $G[F]$.

Given G and $G[F]$, let us recapitulate how we can determine the unknown set A, where $|A| \leq 2$. If there is no edge in $G[F]$, then $A = \emptyset$. If there is only one vertex with degree at least two in $G[F]$, then this vertex is the only element in A ; otherwise, $|A| = 2$. If there are two vertices with degree three in $G[F]$, then these vertices are the two elements in A. If there is no vertex with degree three in $G[F]$, then the elements in A are the two vertices adjacent to vertices of degree one in $G[F]$. Finally, if there is exactly one vertex with degree three in $G[F]$, this vertex belongs to A, and we name it x. Then if there is one vertex which is not adjacent to x in $G[F]$, this vertex is the second element in A ; otherwise, the second element in A is the vertex with degree two in G and in $G[F]$.

Conversely, consider an $(r, \leq 2)$-identifying code $\mathscr{C}'$ of G' with

$$|\mathscr{C}'| \leq k + (6r + 4)m.$$

First we will show that for every $e \in E(G)$ we must have $V(C_e) \subset \mathscr{C}'$. To do this, note that since $\mathscr{C}'$ is an $(r, \leq 2)$-identifying code, if we find in G' two distinct vertices a and b such that

$$(B(a,r) \cup B(b,r)) \, \Delta B(b,r) = \{c\}, \tag{D.4}$$

where $c \in V(G')$ and Δ stands for the symmetric difference of sets, then, since we must have

$$I_r^{\mathscr{C}'}(\{a,b\}) \neq I_r^{\mathscr{C}'}(b),$$

we conclude that necessarily $c \in \mathscr{C}'$. Let us write $(a,b) \longrightarrow c$ if (D.4) is true for three vertices a, b and c. The following facts are easily checked on Fig. D.XIV.

$$(p_{r+1}, p_{r+2}) \longrightarrow p_1$$

$$(p_{r+2}, p_{r+3}) \longrightarrow p_2$$

$$\cdots$$

$$(p_{2r}, p_{2r+1}) \longrightarrow p_r$$

$$(p_{2r+1}, c_1) \longrightarrow p_{r+1}$$

so we must have $p_i \in \mathscr{C}'$ for $i \in \{1, 2, \cdots, r+1\}$. Furthermore

$$(p_2, p_1) \longrightarrow p_{r+2}$$

$$(p_3, p_2) \longrightarrow p_{r+3}$$

$$\cdots$$

$$(p_{r+1}, p_r) \longrightarrow p_{2r+1}$$

and so we also have $p_i \in \mathscr{C}'$ for $i \in \{r+2, r+3, \cdots, 2r+1\}$. The same argument can be applied on the cycle :

$$(c_{r+1}, c_{r+2}) \longrightarrow c_1$$

$$(c_{r+2}, c_{r+3}) \longrightarrow c_2$$

$$\cdots$$

$$(c_{4r+2}, c_{4r+3}) \longrightarrow c_{3r+2}$$

and in the other direction

$$(c_{2r+3}, c_{2r+2}) \longrightarrow c_{3r+3}$$

$$\cdots$$

$$(c_{3r+3}, c_{3r+2}) \longrightarrow c_{4r+3}.$$

Thus for every edge $e \in E(G)$ we have $V(C_e) \subset \mathscr{C}'$. Since $|V(C_e)| = 6r+4$, if we define $\mathscr{C}$ as the trace of $\mathscr{C}'$ on $V(G)$

$$\mathscr{C} = \mathscr{C}' \cap V(G),$$

the cardinal of $\mathscr{C}$ is at most

$$|\mathscr{C}| \leq |\mathscr{C}'| - (6r+4)m \leq k.$$

To conclude, note that for every edge $xy \in E(G)$, we have in C_e

$$B(p_{r+1}, r) \Delta \left(B(p_r, r) \cup B(p_{r+1}, r) \right) = \{x, y\}$$

and so we must have $x \in \mathscr{C}$ or $y \in \mathscr{C}$; thus $\mathscr{C}$ is a vertex cover of G. This ends the proof of Proposition D.10. □

In conclusion, Propositions D.2–D.5 together with Proposition D.10 cover all the cases in Theorem D.1 and give a complete proof for it : we have therefore shown that the problem MIN $(r, \leq l)$-ID-CODE IN Π_3 is NP-complete for $l \in \{1, 2\}$ and all $r \geq 1$.

Annexe E

Minimal Identifying Codes in Trees and Planar Graphs with Large Girth

David Auger [1]
david.auger@telecom-paristech.fr

Abstract

LLet G be a finite undirected graph with vertex set $V(G)$. If $v \in V(G)$, let $N[v]$ denote the closed neighbourhood of v, i.e. v itself and all its adjacent vertices in G. An identifying code in G is a subset $\mathscr{C}$ of $V(G)$ such that the sets $N[v] \cap \mathscr{C}$ are nonempty and pairwise distinct for each vertex $v \in V(G)$. We consider the problem of finding the minimum size of an identifying code in a given graph, which is known to be NP-hard. We give a linear algorithm that solves it in the class of trees, but show that the problem remains NP-hard in the class of planar graphs with arbitrarily large girth.

Keywords : Identifying Codes, Graph Theory,
Trees, NP-complete problems, Complexity, Algorithms.

E.1 Introduction

By a *graph* we mean a finite, undirected graph, without loops nor multiple edges. If G is a graph, we denote respectively by $V(G)$ and $E(G)$ the sets of vertices and edges of G. An edge $\{x, y\} \in E(G)$ with $x, y \in V(G)$ will be simply denoted by xy. We refer to [21] for basic notions such as *adjacent* vertices, *paths*, *cycles* or *connectivity*. Let us just recall that a *tree* is a connected graph with no cycles, and that the *girth* of a graph G is the smallest possible length of a cycle in G.

1. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

The *closed neighbourhood* of a vertex $v \in V(G)$ is the set $N_G[v]$ (or simply $N[v]$ when there is no ambiguity) containing v and all its adjacent vertices in G. We recall that the *degree* of a vertex $v \in V(G)$ is the number $d(v)$ of vertices $w \in V(G)$ such that $vw \in E(G)$. The *maximum degree* of G is the maximum possible degree of a vertex in G. Finally, a graph is *planar* if it can be drawn in the plane in such a way that its edges do not cross. For precise definitions and additional background about graphs, we refer once again to [21]. The reader will also need basic knowledge in algorithmic complexity such as polynomial reduction and NP-completeness ; for these notions we refer to [50].

In this paper, what we call a *code* is simply a set of vertices $\mathscr{C} \subseteq V(G)$, and we refer to its elements as *codewords*. A codeword c in a code $\mathscr{C}$ is said to *cover* a vertex v if $v \in N[c]$; we say that v is covered by $\mathscr{C}$ if it is covered by at least one codeword. If v, w are distinct vertices in $V(G)$, we say that a codeword $c \in \mathscr{C}$ *separates* v and w, or v from w, if c covers exactly one of these two vertices.

An *identifying code* in G is a code $\mathscr{C} \subseteq V(G)$ such that all the sets

$$N[v] \cap \mathscr{C}$$

for $v \in V(G)$ are nonempty and pairwise distinct.

Equivalently, $\mathscr{C}$ is an identifying code if every vertex v of G is *identified* by $\mathscr{C}$, i.e. if v is covered by $\mathscr{C}$ and separated from every other vertex of G by at least one codeword.

The notion of identifying code was introduced in [65] with the original motivation of fault diagnosis in multiprocessor systems. The general idea is the following : assume that vertices in the graph are processors, and that a codeword is a processor equipped with a sensor, with the ability to detect a faulty processor if it is in its closed neighbourhood. Then if there as most one faulty processor and if every codeword sends us one bit of information, refering to whether it detects a faulty processor or not, the fact that $\mathscr{C}$ is an identifying code enable us to know, from the the $|\mathscr{C}|$ bits of information received, if there is a fault in the graph, and also to deduce its position if there is one.

It was proved in [34] that the problem of finding the minimum size of an identifying code in a given graph is NP-hard. In the next sections, we first provide a linear algorithm which computes the minimum size of an identifying code in a given tree, and then exhibit some classes of graphs, in a certain sense as close as desired to the class of trees, where the problem of finding the minimum size of an identifying code remains NP-hard. For combinatorial results about minimal identifying codes in trees, see [18] and [19]. One can also find results for identifying codes in some graphs with high girth in [68].

Algorithms for similar coding problems in trees are known : see [27] for an algorithm computing a mininimal identifying code in an *oriented* tree, and [41] and [86] for algorithms computing minimal locating-dominating codes.

For a comprehensive bibliography about identifying codes and locating-dominating codes, see [73].

E.2 An algorithm for minimum identifying codes in trees

In this section we prove the following result :

Theorem E.1. *There exists a linear algorithm which computes the minimum size of an identifying code in a given tree.*

E.2.1 Almost identifying codes

For the algorithm mentioned in Theorem E.1 we will need the following notion : if G is a graph and $A \subseteq V(G)$, we say that a subset $\mathscr{C}$ of $V(G)$ is an A-*almost identifying code* of G if the sets

$$\mathscr{C} \cap N[v]$$

are all nonempty and pairwise distinct for all v in $V(G) \setminus A$. Thus an $\emptyset$-almost identifying code is just an identifying code.

In an A-almost identifying code, the vertices in A can be chosen as codewords and thus used to identify vertices in $V(G) \setminus A$, but do not need to be identified. If $v \in V(G)$, we write v-*almost identifying* for $\{v\}$-almost identifying.

Consider a graph G, a vertex $v \in V(G)$ and a v-almost identifying code $\mathscr{C}$. We say that

- $\mathscr{C}$ satisfies the ID property (for *identifying*) if $\mathscr{C}$ is an identifying code in G ;
- $\mathscr{C}$ satisfies the CO property (for *code*) if $v \in \mathscr{C}$;
- $\mathscr{C}$ satisfies the ADJ property (for *adjacent*) if v is adjacent to a codeword ;
- $\mathscr{C}$ satisfies the FN property (for *favoured neighbour*) if there exists a neighbour w of v such that $N[w] \cap \mathscr{C} = \{v\}$; in this case we say that w is the favoured neighbour of v, in the sense that v is the only codeword covering w ; since $\mathscr{C}$ is v-almost identifying v admits at most one favoured neighbour ;
- $\mathscr{C}$ satisfies $\overline{\text{ID}}$, $\overline{\text{CO}}$, $\overline{\text{ADJ}}$ or $\overline{\text{FN}}$ respectively if $\mathscr{C}$ does not satisfy properties ID, CO, ADJ or FN.

There exist dependence relationships between these properties, for instance the reader will easily check that :

- if $\mathscr{C}$ satisfies FN, then it satisfies CO ;
- if $\mathscr{C}$ satisfies ID, then it satisfies CO or ADJ ;
- if $\mathscr{C}$ satisfies ID, CO and FN, then it must satisfy ADJ.

E.2.2 Main and auxiliary functions

Let a tree T be given and let $v \in V(T)$. An identifying code of T can be viewed as a v-almost identifying code in T satisfying property ID ; we denote by

$$f_{\text{ID}}(v, T)$$

the minimum size of such a code. More generally, if Π_i denotes a possible property of a v-almost identifying code for $1 \leq i \leq k$ (like ID, $\overline{\text{CO}}$, etc.), we denote by

$$f_{\Pi_1,\cdots,\Pi_k}(v, T)$$

the minimum size of a v-almost identifying code in T satisfying all the properties Π_i for $1 \leq i \leq k$; if such a code does not exist, the function takes the value $+\infty$.

We need to consider 17 functions, the first 10 we call *main functions* and the later 7 *auxiliary functions*. Table E.I gives the list of the main functions, whereas auxiliary functions are given on Table E.II ; this table also gives simple formulas showing how auxiliary functions can be computed from main functions.

number	function
1	$f_{\mathrm{ID,CO,ADJ,FN}}$
2	$f_{\mathrm{ID,CO,ADJ,\overline{FN}}}$
3	$f_{\mathrm{ID,CO,\overline{ADJ}}}$
4	$f_{\mathrm{ID,\overline{CO},ADJ}}$
5	$f_{\mathrm{CO,ADJ,FN}}$
6	$f_{\mathrm{CO,ADJ,\overline{FN}}}$
7	$f_{\mathrm{CO,\overline{ADJ},FN}}$
8	$f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}$
9	$f_{\mathrm{\overline{CO},ADJ}}$
10	$f_{\mathrm{\overline{CO},\overline{ADJ}}}$

Table E.I – *List of main functions*

number	function
11	$f_{\mathrm{CO,\overline{ADJ}}}(v,T) = \min \begin{cases} f_{\mathrm{CO,\overline{ADJ},FN}}(v,T), \\ f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v,T) \end{cases}$
12	$f_{\mathrm{CO,ADJ}}(v,T) = \min \begin{cases} f_{\mathrm{CO,ADJ,FN}}(v,T), \\ f_{\mathrm{CO,ADJ,\overline{FN}}}(v,T) \end{cases}$
13	$f_{\mathrm{CO,FN}}(v,T) = \min \begin{cases} f_{\mathrm{CO,ADJ,FN}}(v,T), \\ f_{\mathrm{CO,\overline{ADJ},FN}}(v,T) \end{cases}$
14	$f_{\mathrm{CO,\overline{FN}}}(v,T) = \min \begin{cases} f_{\mathrm{CO,ADJ,\overline{FN}}}(v,T), \\ f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v,T) \end{cases}$
15	$f_{\mathrm{\overline{CO}}}(v,T) = \min \begin{cases} f_{\mathrm{\overline{CO},ADJ}}(v,T), \\ f_{\mathrm{\overline{CO},\overline{ADJ}}}(v,T) \end{cases}$
16	$f_{\mathrm{CO}}(v,T) = \min \begin{cases} f_{\mathrm{CO,ADJ,FN}}(v,T), \\ f_{\mathrm{CO,ADJ,\overline{FN}}}(v,T), \\ f_{\mathrm{CO,\overline{ADJ},FN}}(v,T), \\ f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v,T) \end{cases}$
17	$f_{\mathrm{ID,CO}}(v,T) = \min \begin{cases} f_{\mathrm{ID,CO,ADJ,FN}}(v,T), \\ f_{\mathrm{ID,CO,ADJ,\overline{FN}}}(v,T), \\ f_{\mathrm{ID,CO,\overline{ADJ}}}(v,T) \end{cases}$

Table E.II – *List of auxiliary functions*

E.2.3 The algorithm *AIC*

The algorithm mentionned in Theorem E.1 consists in choosing (randomly) a vertex v_1 in a given graph T and computing the values of the 17 main and auxiliary functions on (v_1, T), and then computing $f_{\text{ID}}(v_1, T)$ by

$$f_{\text{ID}}(v, T) = \min \begin{cases} f_{\text{ID,CO,ADJ,FN}}(v_1, T), \\ f_{\text{ID,CO,ADJ},\overline{\text{FN}}}(v_1, T), \\ f_{\text{ID,CO},\overline{\text{ADJ}}}(v_1, T), \\ f_{\text{ID},\overline{\text{CO}},\text{ADJ}}(v_1, T). \end{cases}$$

Remember that this value is the minimum size of an identifying code in T. In order to do this we define an algorithm AIC (for *almost identifying code*), which returns the values of the 17 main and auxiliary functions for a given pair (v_1, T), where T is a tree and v_1 a vertex of T.

Algorithm 1 AIC

Input : a tree T and a vertex v_1 of T.
Output : the list ℓ of the values of the 17 main and auxiliary functions on (v_1, T).

1: **if** v_1 has degree 0 in T **then**
2: initalize ℓ (Table E.III) ;
3: **else**
4: let v_2 be a neighbour of v_1 in T ;
5: let T_1 and T_2 be the trees, respectively containing v_1 and v_2 as vertices, obtained from T by deletion of the edge v_1v_2 ;
6: let l_1 =AIC(v_1, T_1) and l_2 =AIC(v_2, T_2) ;
7: compute the 10 main functions on (v_1, T) from l_1 and l_2 (Table E.IV) ;
8: compute the 7 auxiliary functions on (v_1, T) from the main functions on (v_1, T) (Table E.II) ;
9: **end if**
10: return the list ℓ of the values the 17 main and auxiliary functions on (v, T).

Algorithm AIC recursively computes the values of the 17 functions in smaller and smaller trees. It uses the following facts :

– first, if T consists of a single vertex v_1, then the values of the 17 functions are easy to compute. These values are given on Table E.III (used on line 2 of the algorithm) ;
– second, one can compute the values of the 7 auxiliary functions on (v_1, T) from the values of the 10 main functions on (v_1, T), by the formulas given on Table E.II (used on line 8 of the algorithm) ;
– finally, if T consists of at least two vertices, then the vertex v_1 has at least one neighbour v_2 in T ; this is the central step of the algorithm where we use recursion. If we remove the edge v_1v_2 from T, we obtain two trees T_1 and T_2 respectively containing v_1 and v_2 (see Fig. E.I). We claim that the values of the 10 main functions on (v_1, T) can be computed from the values of the 17 main and auxiliary functions on (v_1, T_1) and (v_2, T_2). The formulas showing how this can be done are given in Table E.IV. This is used on line 7 of the algorithm.

The first two facts are easy to check, only the last one needs a proof. Since proving all cases in Table E.IV could be long and tedious, we give hereafter a detailed proof of the first formula as a corollary to Lemma E.2 ; the proof of all other cases is similar. However, we give in the appendix an exhaustive list of figures that can be used to check all cases.

number	function	value	number	function	value
1	$f_{\mathrm{ID,CO,ADJ,FN}}$	$+\infty$	10	$f_{\overline{\mathrm{CO}},\overline{\mathrm{ADJ}}}$	0
2	$f_{\mathrm{ID,CO,ADJ},\overline{\mathrm{FN}}}$	$+\infty$	11	$f_{\mathrm{CO},\overline{\mathrm{ADJ}}}$	1
3	$f_{\mathrm{ID,CO},\overline{\mathrm{ADJ}}}$	1	12	$f_{\mathrm{CO,ADJ}}$	$+\infty$
4	$f_{\mathrm{ID},\overline{\mathrm{CO}},\mathrm{ADJ}}$	$+\infty$	13	$f_{\mathrm{CO,FN}}$	$+\infty$
5	$f_{\mathrm{CO,ADJ,FN}}$	$+\infty$	14	$f_{\mathrm{CO},\overline{\mathrm{FN}}}$	1
6	$f_{\mathrm{CO,ADJ},\overline{\mathrm{FN}}}$	$+\infty$	15	$f_{\overline{\mathrm{CO}}}$	0
7	$f_{\mathrm{CO},\overline{\mathrm{ADJ}},\mathrm{FN}}$	$+\infty$	16	f_{CO}	1
8	$f_{\mathrm{CO},\overline{\mathrm{ADJ}},\overline{\mathrm{FN}}}$	1	17	$f_{\mathrm{ID,CO}}$	1
9	$f_{\overline{\mathrm{CO}},\mathrm{ADJ}}$	$+\infty$			

Table E.III – *Initialization of the* 17 *functions on a single vertex*

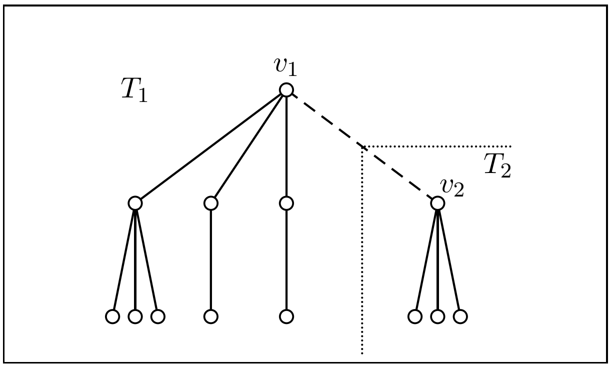


Figure E.I – *The trees T_1 and T_2 resulting from the deletion of the edge v_1v_2*

Lemma E.2. *Let T be a tree and v_1v_2 be an edge of T. Let T_1 and T_2 be the trees, respectively containing v_1 and v_2 as vertices, obtained from T by deletion of the edge v_1v_2. Let $\mathscr{C}$ be a code in T and $\mathscr{C}_i = \mathscr{C} \cap V(T_i)$ for $i \in \{1, 2\}$. Then $\mathscr{C}$ is a v_1-almost identifying code in T with properties* ID, CO, ADJ, FN *if and only if $\mathscr{C}_1$ is a v_1-almost identifying code in T_1 and $\mathscr{C}_2$ is a v_2-almost identifying code in T_2, and one of the following assertions is satisfied :*

1. *$\mathscr{C}_1$ satisfies* ID, CO, ADJ, $\overline{\mathrm{FN}}$ *and $\mathscr{C}_2$ satisfies* $\overline{\mathrm{CO}}, \overline{\mathrm{ADJ}}$ *;*
2. *$\mathscr{C}_1$ satisfies* CO, ADJ, FN *and $\mathscr{C}_2$ satisfies* CO, $\overline{\mathrm{ADJ}}$ *;*
3. *$\mathscr{C}_1$ satisfies* ID, CO, ADJ, FN *and $\mathscr{C}_2$ satisfies* $\overline{\mathrm{CO}}$, ADJ *;*
4. *$\mathscr{C}_1$ satisfies* CO, FN *and $\mathscr{C}_2$ satisfies* CO, ADJ*.*

Proof. This case is depicted on Fig. I in the appendix.

Suppose first that $\mathscr{C}$ is a v_1-almost identifying code in T with properties ID, CO, ADJ, FN.

Observe that when going from T to T_1, only v_1 in T_1 loses a neighbour in the operation and so in T_1, for $v \neq v_1$, we have $N_{T_1}[v] \cap \mathscr{C}_1 = N_T[v] \cap \mathscr{C}$, thus $\mathscr{C}_1$ is a v_1-almost identifying code in T_1 ; a similar argument shows that $\mathscr{C}_2$ is a v_2-almost identifying code in T_2.

Next consider the code $\mathscr{C}_2$; elementary logic implies that one of the four possibilites ($\overline{\mathrm{CO}}$ and $\overline{\mathrm{ADJ}}$), (CO and $\overline{\mathrm{ADJ}}$), ($\overline{\mathrm{CO}}$ and ADJ) or (CO and ADJ) must happen. In all cases, since $\mathscr{C}$ satisfies CO and v_1 is a vertex of T_1, the code $\mathscr{C}_1$ will satisfy CO.

If $\mathscr{C}_2$ satisfies $\overline{\mathrm{CO}}, \overline{\mathrm{ADJ}}$ (top-left square on Fig. I), since $v_2 \notin \mathscr{C}_2$ and $\mathscr{C}$ satisfies ADJ, the code $\mathscr{C}_1$ must satisfy ADJ. Since $v_2 \notin \mathscr{C}$, it cannot contribute to identify v_1 in T and so $\mathscr{C}_1$

must satisfy ID. Finally, since $\mathscr{C}$ is v_1-almost identifying and v_2 is a favoured neighbour of v_1, no favoured neighbour for v_1 is allowed in T_1 : thus $\mathscr{C}_1$ satisfies ID, CO, ADJ and $\overline{\text{FN}}$.

If $\mathscr{C}_2$ satisfies CO, $\overline{\text{ADJ}}$ (top-right square on Fig. I), since $\mathscr{C}$ satisfies ID, the vertices v_1 and v_2 must be separated by a codeword of $\mathscr{C}$, which is either a neighbour of v_1, distinct from v_2, or a neighbour of v_2, distinct from v_1 ; but since $\mathscr{C}_2$ satisfies $\overline{\text{ADJ}}$ the second possibility cannot happen and so $\mathscr{C}_1$ satisfies ADJ. Next, since $\mathscr{C}$ satisfies FN and $\mathscr{C}_2$ satisfies CO, the favored neighbour of v_1 is not v_2, and so $\mathscr{C}_1$ must satisfy FN. Thus $\mathscr{C}_1$ satisfies CO, ADJ and FN.

If $\mathscr{C}_2$ satisfies $\overline{\text{CO}}$, ADJ (bottom-left square on Fig. I), since v_2 cannot contribute to the identification of v_1 by $\mathscr{C}$ in T, $\mathscr{C}_1$ must satisfy ID. Since $v_2 \notin \mathscr{C}$ and $\mathscr{C}$ satisfies ADJ, this must also be the case for $\mathscr{C}_1$. Finally, by ADJ for $\mathscr{C}_2$, the favored neighbour of v_1 in T is not v_2 and so $\mathscr{C}_1$ satisfies FN. Thus $\mathscr{C}_1$ satisfies ID, CO, ADJ and FN.

Eventually, if $\mathscr{C}_2$ satisfies CO, ADJ (bottom-right square on Fig. I), then once again the favoured neighbour of v_1 in T must be found in T_1 and $\mathscr{C}_1$ satisfies FN. Thus $\mathscr{C}_1$ satisfies CO and FN.

For the converse implications, let us consider the first case : suppose that $\mathscr{C}_1$ satisfies ID, CO, ADJ, $\overline{\text{FN}}$ and $\mathscr{C}_2$ satisfies $\overline{\text{CO}}$, $\overline{\text{ADJ}}$, and let us define $\mathscr{C} = \mathscr{C}_1 \cup \mathscr{C}_2$, a code in T (see the top-left square on Fig. I) :

- if $v \in V(T_1)$ we have $N_T[v] \cap \mathscr{C} = N_{T_1}[v] \cap \mathscr{C}_1$, and so all vertices in $V(T_1)$ are covered by nonempty and distinct sets of codewords from $\mathscr{C}$ (including v_1, since $\mathscr{C}_1$ satisfies ID) ;
- a similar obervation can be made for vertices in $V(T_2) \setminus \{v_2\}$;
- if $v \in V(T_1)$ and $v' \in V(T_2) \setminus \{v_2\}$, then $N_T[v] \cap N_T[v']$ is empty or equal to $\{v_2\}$, but since $v_2 \notin \mathscr{C}$ we conclude that vertices in $V(T_1)$ are separated from vertices in $V(T_2) \setminus \{v_2\}$: a codeword covering v cannot cover v' ;
- it remains to settle the score for v_2 : we have $N_T[v_2] \cap \mathscr{C} = \{v_1\}$, so v_2 is covered by $\mathscr{C}$; it is separated from v_1, since $\mathscr{C}_1$ satisfies ADJ, and separated from all vertices in $V(T_1) \setminus \{v_1\}$ since these vertices cannot be covered only by v_1 ($\mathscr{C}_1$ satisfies $\overline{\text{FN}}$). Finally, v_2 is separated from vertices in $V(T_2) \setminus \{v_2\}$ since these vertices cannot be covered by v_1.

We have proved that $\mathscr{C}$ is an identifying code, i.e. a v_1-almost identifying code satisfying ID. Moreover, $\mathscr{C}$ obviously satisfies CO, ADJ and FN, with v_2 as the favoured neighbour of v_1.

We skip the proofs for the three remaining cases which are very much similar to the previous one. □

From this lemma we directly deduce the following equality :

Corollary E.3. *With the notation of Lemma E.2, we have the following equality :*

$$f_{\text{ID,CO,ADJ,FN}}(v_1, T) = \min \begin{cases} f_{\text{ID,CO,ADJ},\overline{\text{FN}}}(v_1, T_1) + f_{\overline{\text{CO}},\overline{\text{ADJ}}}(v_2, T_2), \\ f_{\text{CO,ADJ,FN}}(v_1, T_1) + f_{\text{CO},\overline{\text{ADJ}}}(v_2, T_2), \\ f_{\text{ID,CO,ADJ,FN}}(v_1, T_1) + f_{\overline{\text{CO}},\text{ADJ}}(v_2, T_2), \\ f_{\text{CO,FN}}(v_1, T_1) + f_{\text{CO,ADJ}}(v_2, T_2). \end{cases}$$

With the help of the appendix, an interested reader can easily check all formulas given on Table. E.IV, and conclude that the algorithm is valid. Before ending this part, let us notice that in the execution of the algorithm, when the function AIC(v_1, T) is called, if the degree of v is at least one then an edge is removed from T before computing AIC(v_1, T_1)

1	$f_{\mathrm{ID,CO,ADJ,FN}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{ID,CO,ADJ,\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{CO,ADJ,FN}}(v_1,T_1)+f_{\mathrm{CO,\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{ID,CO,ADJ,FN}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2), \\ f_{\mathrm{CO,FN}}(v_1,T_1)+f_{\mathrm{CO,ADJ}}(v_2,T_2) \end{cases}$
2	$f_{\mathrm{ID,CO,ADJ,\overline{FN}}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{CO,ADJ,\overline{FN}}}(v_1,T_1)+f_{\mathrm{CO,\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{ID,CO,ADJ,\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2), \\ f_{\mathrm{CO,\overline{FN}}}(v_1,T_1)+f_{\mathrm{CO,ADJ}}(v_2,T_2) \end{cases}$
3	$f_{\mathrm{ID,CO,\overline{ADJ}}}(v_1,T)$	$=$	$f_{\mathrm{ID,CO,\overline{ADJ}}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2)$
4	$f_{\mathrm{ID,\overline{CO},ADJ}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{\overline{CO},ADJ}}(v_1,T_1)+f_{\mathrm{ID,CO,\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{ID,\overline{CO},ADJ}}(v_1,T_1)+f_{\mathrm{ID,\overline{CO},ADJ}}(v_2,T_2), \\ f_{\mathrm{\overline{CO},ADJ}}(v_1,T_1)+f_{\mathrm{ID,CO,ADJ,FN}}(v_2,T_2), \\ f_{\mathrm{\overline{CO}}}(v_1,T_1)+f_{\mathrm{ID,CO,ADJ,\overline{FN}}}(v_2,T_2) \end{cases}$
5	$f_{\mathrm{CO,ADJ,FN}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{CO,ADJ,\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{CO,ADJ,FN}}(v_1,T_1)+f_{\mathrm{CO,\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{CO,ADJ,FN}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2), \\ f_{\mathrm{CO,FN}}(v_1,T_1)+f_{\mathrm{CO,ADJ}}(v_2,T_2) \end{cases}$
6	$f_{\mathrm{CO,ADJ,\overline{FN}}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{CO,\overline{FN}}}(v_1,T_1)+f_{\mathrm{CO}}(v_2,T_2), \\ f_{\mathrm{CO,ADJ,\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2) \end{cases}$
7	$f_{\mathrm{CO,\overline{ADJ},FN}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},\overline{ADJ}}}(v_2,T_2), \\ f_{\mathrm{CO,\overline{ADJ},FN}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2) \end{cases}$
8	$f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v_1,T)$	$=$	$f_{\mathrm{CO,\overline{ADJ},\overline{FN}}}(v_1,T_1)+f_{\mathrm{\overline{CO},ADJ}}(v_2,T_2)$
9	$f_{\mathrm{\overline{CO},ADJ}}(v_1,T)$	$=$ min	$\begin{cases} f_{\mathrm{\overline{CO}}}(v_1,T_1)+f_{\mathrm{ID,CO}}(v_2,T_2), \\ f_{\mathrm{\overline{CO},ADJ}}(v_1,T_1)+f_{\mathrm{ID,\overline{CO},ADJ}}(v_2,T_2) \end{cases}$
10	$f_{\mathrm{\overline{CO},\overline{ADJ}}}(v_1,T)$	$=$	$f_{\mathrm{\overline{CO},\overline{ADJ}}}(v_1,T_1)+f_{\mathrm{ID,\overline{CO},ADJ}}(v_2,T_2)$

Table E.IV – *Recurrence formulas for main functions* and AIC(v_2, T_2), and thus the number of calls to the function AIC is at most the number of edges in T. Since each step can be executed in constant time, we conclude that the algorithm runs in linear time. Let us also note that this algorithm could be easily modified in order to output an identifying code of minimal size (it would be sufficient in each computation to keep track of the functions which give the minimal values), or else to compute the number of identifying codes with minimal size in T.

E.3 Planar graphs with large girth

We proved in the previous section that the problem of finding the minimum size of an identifying code can be solved in linear time in the class of trees. Since it is known that the problem is NP-hard in the general case (see [34]), we found interesting to narrow the gap between these two extremes. Without loss of generality, we restrict ourselves to connected graphs. If $\mathcal{H}$ is a class of graphs, let us call Min ID-code in $\mathcal{H}$ the problem of deciding, for a given graph $G \in \mathcal{H}$ and an integer p, whether G admits an identifying code of size at most p or not.

Let $\mathcal{P}_k^4$ denote the class of connected planar graphs, with maximum degree at most 4, and girth at least k where $k \geq 3$. It should be noted that if k is large the elements of $\mathcal{P}_k^4$ are "nearly" trees, in the sense that

$$\bigcap_{k \geq 3} \mathcal{P}_k^4$$

is the class of trees with maximum degree 4.

We prove the following result :

Theorem E.4. *For all $k \geq 3$, the problem* Min ID-code in $\mathcal{P}_k^4$ *is NP-complete.*

Let us start by a lemma.

Lemma E.5. *Let $P = av_1v_2 \cdots v_{2k}b$ be a path on $2k+2$ vertices, where $k \geq 1$. Then :*
- *the minimal size of an $\{a, b\}$-almost identifying code in P which contains neither a nor b is $k+1$;*
- *the minimal size of an $\{a, b\}$-almost identifying code in P which contains exactly one of a and b is $k+1$;*
- *the minimal size of an $\{a, b\}$-almost identifying code in P which contains a and b is $k+2$.*

Proof.

Let $\mathscr{C}$ be an $\{a, b\}$-almost identifying code in P. For every i such that $1 \leq i \leq 2k-1$, the vertices v_i and v_{i+1} must be separated by a codeword ; let us consider this as a *task* that has to be fulfilled. The vertices v_1 and v_{2k} must be covered by $\mathscr{C}$, giving us 2 other tasks, for a total of $2k+1$ tasks. Suppose now that v_i is a codeword : if $i \in \{3, \cdots, 2k-2\}$, it covers neither v_1 nor v_{2k}, but it separates v_{i-1} from v_{i-2} and v_{i+1} from v_{i+2} ; thus v_i fulfills exactly 2 tasks. If $i \in \{1, 2\}$, then v_i covers v_1 but only separates the vertices v_{i+1} and v_{i+2}, thus also fulfills 2 tasks, and a similar observation can be made if $i \in \{2k-1, 2k\}$, and for the vertices a and b.

Therefore, since we have $2k+1$ tasks and since a given codeword fulfills exactly two of them, we need at least $\left\lceil \frac{2k+1}{2} \right\rceil = k+1$ codewords in $\mathscr{C}$. Now since $\mathscr{C} \cap \{v_1, v_2, \cdots, v_{2k}\}$ is a $\{v_1, v_{2k}\}$-almost identifying code in the path $P' = v_1v_2 \cdots v_{2k}$, the same observation leads to conclusion that there are at least k codewords in $\{v_1, v_2, \cdots, v_{2k}\}$: so if at most one of the vertices a, b is a codeword, we have $|\mathscr{C}| \geq k+1$, and if a and b are codewords we have $|\mathscr{C}| \geq k+2$.

Conversely, it is easy to see that the codes

$$\mathscr{C}_1 = \{v_2, v_4, \cdots, v_{2k}\} \cup \{v_3\},$$
$$\mathscr{C}_2 = \{a, v_2, v_4, \cdots, v_{2k}\}$$

and

$$\mathscr{C}_3 = \{a, b\} \cup \{v_2, v_4, \cdots, v_{2k}\}$$

are $\{a, b\}$-almost identifying with the required conditions. □

Before proving Theorem E.4, let us note that one can rapidly check if a given code in a graph is identifying, and so the problem Min ID-code in $\mathcal{P}_k^4$ is in the class NP for all $k \geq 3$. In order to prove its NP-completeness it remains to polynomially reduce an NP-complete problem to Min ID-code in $\mathcal{P}_k^4$. We use the Min Vertex Cover in $\mathcal{P}^3$ problem.

Let $\mathcal{P}^3$ denote the class of planar graphs with maximum degree at most 3. We recall that a *vertex cover* in a graph G is a code $\mathscr{C} \subseteq V(G)$ such that for every edge $e = ab \in E(G)$, one has $a \in \mathscr{C}$ or $a \in \mathscr{C}$ (or both). The following problem was proved to be NP-complete in [49] :

Min Vertex Cover in $\mathcal{P}^3$

- Instance : a planar graph $G \in \mathcal{P}^3$, and an integer p ;
- Question : is there a vertex cover $\mathscr{C}$ of G with $|\mathscr{C}| \leq p$?

Thanks to this result we can now prove Theorem E.4.

Proof of theorem E.4.

Let $k \geq 3$. Let $G \in \mathcal{P}^3$ and $p \geq 0$ be an instance of Min Vertex Cover in $\mathcal{P}^3$; let n and m respectively denote the number of vertices and edges in G. We give a polynomial time construction of a graph $G' \in \mathcal{P}_k^4$ such that

G admits a vertex cover of size at most p if and only if G' admits an identifying code of size at most $p + 3n + km$. (E.1)

This will settle the polynomial reduction and thus prove the theorem. The construction goes as follows : we keep the vertices of G but remove all edges. If two vertices a, b were adjacent in G, via the edge $e = ab$, we link them in G' by a path P_{ab} with $2k$ inner vertices. Finally, we link to every vertex v of G a structure S_v which is depicted on Fig. E.II. It will be convenient, in this construction, to see $V(G)$ as a subset of $V(G')$. An example of transformation for a simple graph is depicted on Fig. E.III.

Obviously, the maximum degree of G' is the maximum degree of G plus one (because of the structures S_v) and G' is planar if G is. We can also note that since edges of G have been replaced by paths of length $2k+1$, the girth of G' is at least $3(2k+1) \geq k$. Thus $G' \in \mathcal{P}_k^4$ if $G \in \mathcal{P}^3$, and the construction is clearly polynomial when k is fixed.

We now prove (E.1). First, assume that $\mathscr{C}$ is a vertex cover of G with size at most p. Since $V(G)$ is a subset of $V(G')$ we can consider $\mathscr{C}$ as a code in G'. Let us start with $\mathscr{C}' := \mathscr{C}$ and add vertices to $\mathscr{C}'$ in order to build an identifying code of G' :

– for $v \in V(G)$, we add to $\mathscr{C}'$ the vertices v_0, v_1 and v_1' in the corresponding structure S_v ;

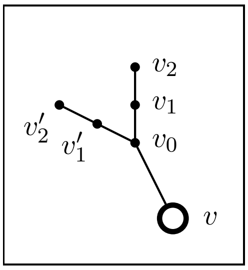


Figure E.II – *The structure S_v linked to the vertex v of G in the proof of Theorem E.4*

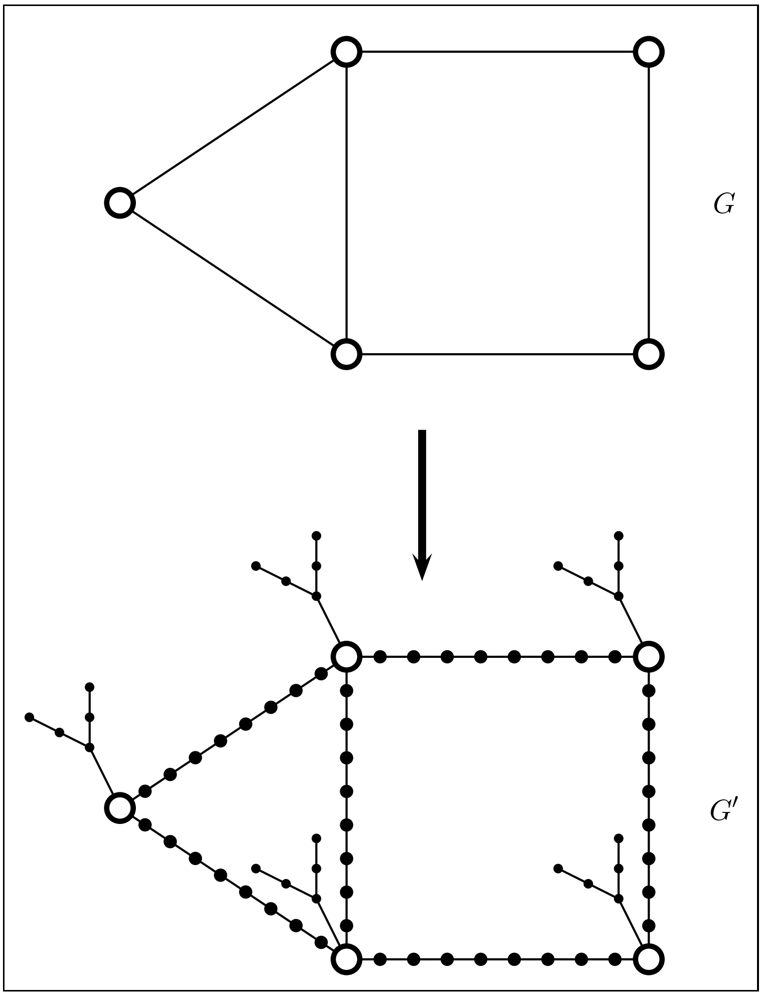


Figure E.III – *An example of transformation for $k = 4$ in the proof of Theorem E.4*

- for every edge ab of G, since $\mathscr{C}$ is a vertex cover of G we must have $a \in \mathscr{C}$ or $b \in \mathscr{C}$. Let us denote by $av_1v_2 \cdots v_{2k}b$ the vertices of the path $aP_{ab}b$, and suppose for instance that $a \in \mathscr{C}$: then we add to $\mathscr{C}'$ the vertices $v_2, v_4, \cdots, v_{2k}$ of P_{ab}.

By doing so, we obtain a code $\mathscr{C}'$ with size

$$|\mathscr{C}'| = |\mathscr{C}| + 3n + km \leq p + 3n + km.$$

It remains to see that $\mathscr{C}'$ is an identifying code of G'. One can easily see that the structures S_v take care of covering and identifying themselves and the vertices of G ; thus we just have to look at what happens in the paths P_{ab} where the conclusion follows as in the proof Lemma E.5. Note in particular that since $\mathscr{C}$ is a vertex cover of G, for every path P_{ab} at least one of the vertices a and b is a codeword.

Conversely, suppose that $\mathscr{C}'$ is an identifying code of G' with size at most $p+3n+km$. Let us recall that the vertices of G' can be partitioned in the following way :

$$V(G') = V(G) \cup \bigcup_{v \in V(G)} V(S_v) \cup \bigcup_{ab \in E(G)} V(P_{ab}).$$

Then :

- for every $v \in V(G)$, consider the vertices of S_v : v_1 and v_2 must be separated, so we must have $v_0 \in \mathscr{C}'$, and v_2, as well as v_2', must be covered, so $v_1 \in \mathscr{C}'$ or $v_2 \in \mathscr{C}'$, and $v_1' \in \mathscr{C}'$ or $v_2' \in \mathscr{C}'$. All in all, there are at least three codewords of $\mathscr{C}'$ in each S_v ;
- for every edge $ab \in E(G)$, by Lemma E.5 the path P_{ab} must count at least $k+1$ codewords if neither a nor b belongs to $\mathscr{C}'$, whereas it must count at least k codewords in the general case.

Let q be the number of *bad* edges of G for the vertex cover, i.e. edges $ab \in E(G)$ such that $a \notin \mathscr{C}'$ and $b \notin \mathscr{C}'$. Then we have

$$\left|\mathscr{C}' \cap V(G)\right| \leq \left|\mathscr{C}'\right| - 3n - (k+1)q - k(m-q)$$

and so since $|\mathscr{C}'| \leq p + 3n + km$ it follows that

$$\left|\mathscr{C}' \cap V(G)\right| \leq p - q.$$

Thus $\mathscr{C}' \cap V(G)$ is a code in G which may not be a vertex cover ; but if we add q vertices to $\mathscr{C}$ (one for every bad edge) we get a vertex cover of G with size at most p. $\square$

Appendix

On the following figures codewords are in black, whereas white vertices are not codewords. An ellipse around some vertices with the mention 'FN' means that one of these vertices is a favoured neighbour of v_1 or v_2.

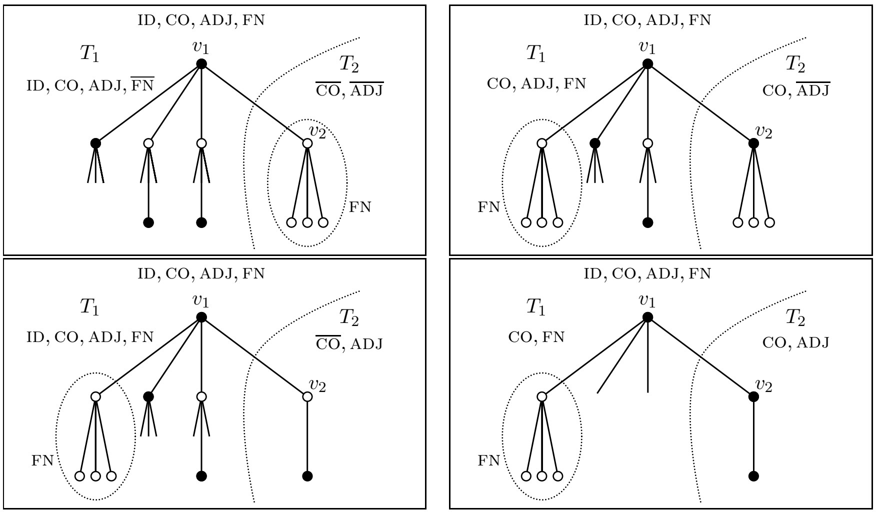


Figure I – *Computation of* $f_{\text{ID,CO,ADJ,FN}}(v_1, T)$ *: 4 cases*

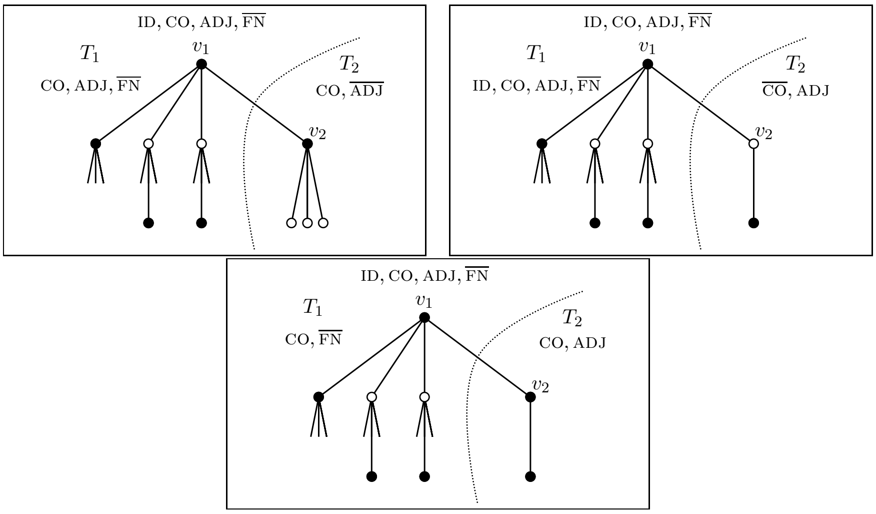


Figure II – *Computation of* $f_{\text{ID,CO,ADJ,}\overline{\text{FN}}}(v_1, T)$ *: 3 cases*

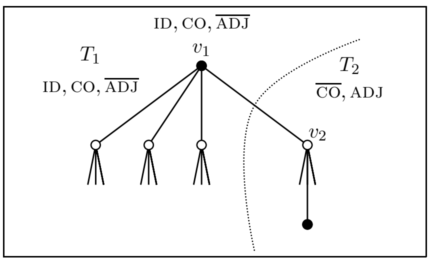


Figure III – *Computation of* $f_{\text{ID,CO,}\overline{\text{ADJ}}}(v_1, T)$ *: 1 case*

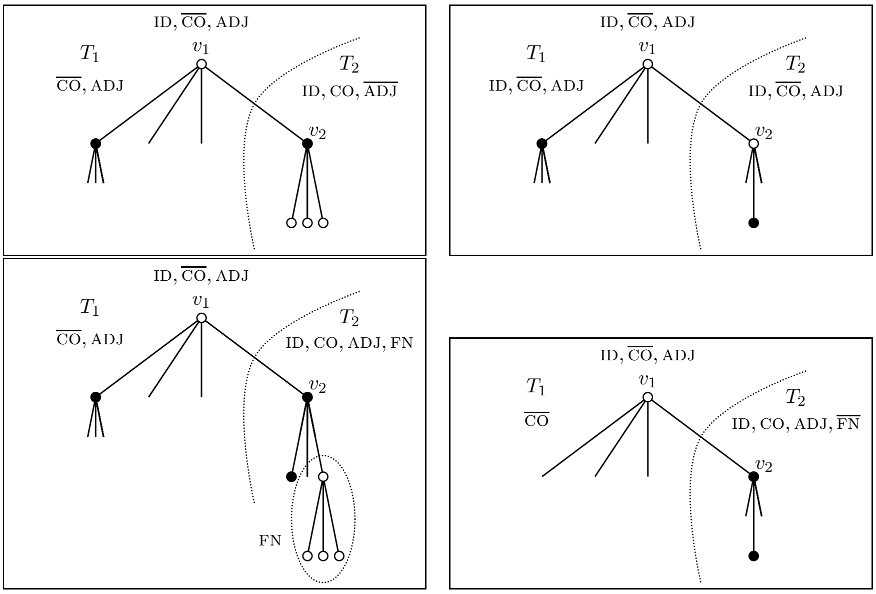


Figure IV – *Computation of* $f_{\mathrm{ID},\overline{\mathrm{CO}},\mathrm{ADJ}}(v_1, T)$ *: 4 cases*

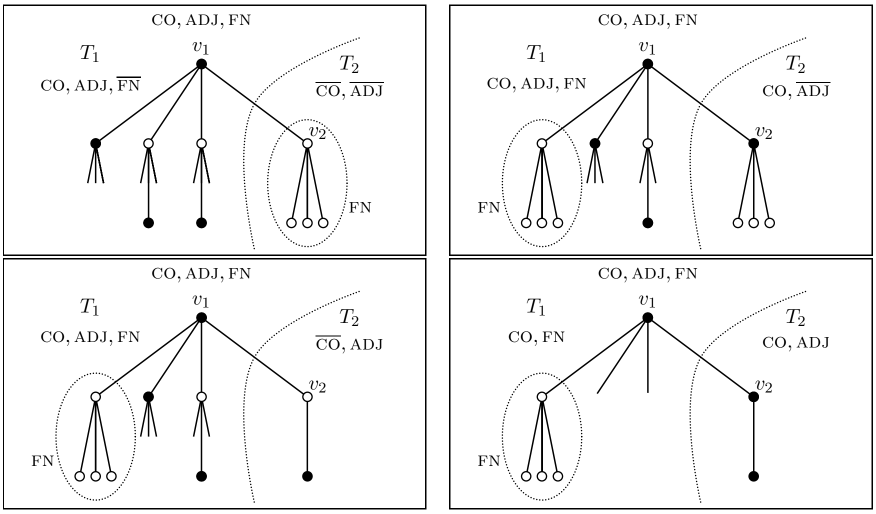


Figure V – *Computation of* $f_{\mathrm{CO},\mathrm{ADJ},\mathrm{FN}}(v_1, T)$ *: 4 cases*

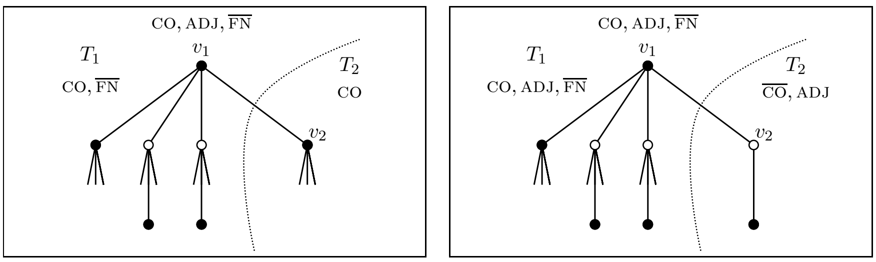


Figure VI – *Computation of* $f_{\mathrm{CO},\mathrm{ADJ},\overline{\mathrm{FN}}}(v_1, T)$ *: 2 cases*

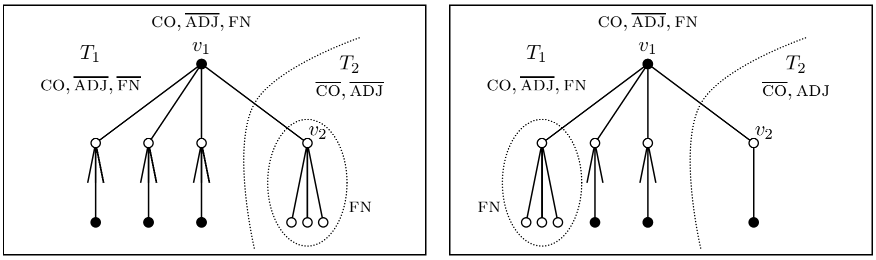


Figure VII – *Computation of* $f_{\text{CO},\overline{\text{ADJ}},\text{FN}}(v_1, T)$ *: 2 cases*

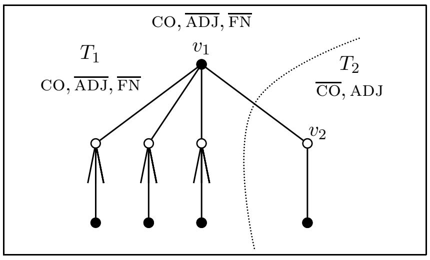


Figure VIII – *Computation of* $f_{\text{CO},\overline{\text{ADJ}},\overline{\text{FN}}}(v_1, T)$ *: 1 case*

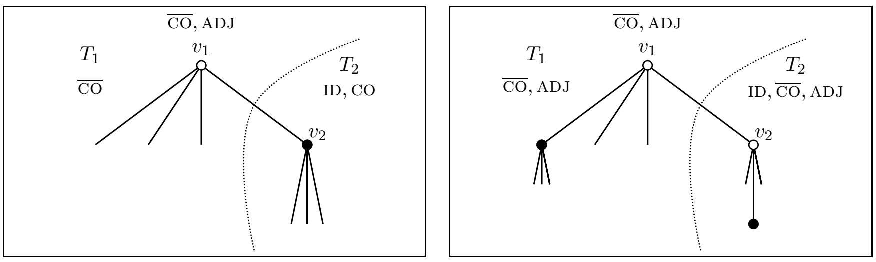


Figure IX – *Computation of* $f_{\overline{\text{CO}},\text{ADJ}}(v_1, T)$ *: 2 cases*

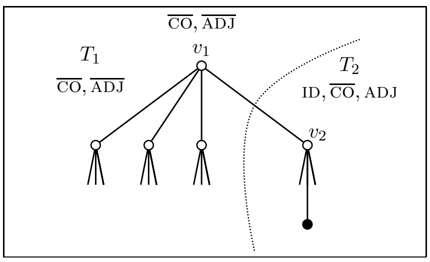


Figure X – *Computation of* $f_{\overline{\text{CO}},\overline{\text{ADJ}}}(v_1, T)$ *: 1 case*

Annexe F

Watching systems in graphs

David Auger*, **Irène Charon***,
Olivier Hudry [1] , **Antoine Lobstein** [2]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Résumé

Nous donnons la définition d'un système de contrôle dans un graphe, qui généralise celle de code identifiant. Nous fournissons quelques propriétés de base, une borne supérieure sur la taille minimum d'un système de contrôle ainsi que des résultats sur les graphes qui atteignent cette borne ; nous étudions aussi les cas de la chaîne et du cycle, et établissons des résultats de complexité.

Abstract

We introduce the notion of watching systems in graphs, which is a generalization of that of identifying codes. We give some basic properties of watching systems, an upper bound on the minimum size of a watching system, and results on the graphs which achieve this bound ; we also study the cases of the paths and cycles, and give complexity results.

Mots-clés : Théorie des graphes, Complexité, Codes identifiants, Systèmes de contrôle, Chaînes, Cycles

Key Words : Graph theory, Complexity, Identifying codes, Watching systems, Paths, Cycles

F.1 Introduction and definitions

F.1.1 Identifying systems

Many search problems, either mathematical problems or 'real life' issues, come down to determine whether a particular item lies in a given set X of possible locations, and locate it if this is the case, by asking questions about its location. Let us suggest a simple model for this, in the so-called *non-adaptive* case, when all questions must be prepared in advance, before getting the answers.

Consider a finite set X and assume that we can only query whether the item lies in certain sets $S \subseteq X$ that belong to a given family $\mathcal{S}$ of subsets of X. For $x \in X$, the *$\mathcal{S}$-identifying set*, or *$\mathcal{S}$-label* (or simply *label* if there is no ambiguity) of x is the set

$$L_{\mathcal{S}}(x) = \{S \in \mathcal{S} : x \in S\}.$$

We say that $\mathcal{S}$ is an *identifying system* of X if the labels of the elements of X are all nonempty and pairwise distinct.

In this case, we can simply ask if the item belongs to S for every $S \in \mathcal{S}$: either all the answers will be negative and the item cannot be in X, or the set of questions with positive answers will correspond to the label $L_{\mathcal{S}}(x)$ of the location x where the item is located. Since $|\mathcal{S}|$ can be much larger than the minimum number of questions required to always succeed, an interesting problem is to find an identifying system $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$ and with minimum size.

Let us mention graph theoretical problems that are particular instances of this general framework. Karpovsky, Chakrabarty and Levitin introduced the notion of *identifying codes* in [65] ; here, with the previous notation, X is the set of vertices of a finite, (in general) undirected graph and $\mathcal{S}$ is the set of all the closed neighbourhoods of the vertices of the graph (see the next section for details). More generally, with the so-called $(r, \leq \ell)$-identifying codes, one can identify sets of vertices within a certain distance (see for instance [70], [69] or [75]). Honkala, Karpovsky and Litsyn, as well as Rosendahl, studied the identification of vertices and edges of a graph using cycles (see [58], [59], [83], [84]). In [62], Honkala and Lobstein considered the identification of vertices in Z^2, using subsets of Z^2. Charbit, Charon, Cohen, Hudry and Lobstein studied the general problem of identifying systems in a bipartite graph framework ([23], [24], [26], [25]). In this paper, we will introduce a new problematics in graphs, which extends the concept of identifying codes, and which can be thought of as identifying vertices with *subsets* of the closed neighbourhoods of the vertices of the graph.

F.1.2 Notation

We use standard notation : by *graph* we mean a simple, finite, undirected, generally connected, graph (if the graph is not connected, we can consider separately its connected components). If G is a graph, we denote its vertex set by $V(G)$ and its edge set by $E(G)$. The *closed neighbourhood* $N_G[v]$ of a vertex v consists of v and its neighbours in G. For $r \geq 0$ and $v \in V(G)$, the *ball of radius r and centre v* is the set $B_G(v, r)$ of all vertices $x \in V(G)$ satisfying $d_G(v, x) \leq r$, where d_G is the usual distance in G. Obviously, $B_G(v, 1) = N_G[v]$. For standard notions such as degree, diameter, spanning tree, etc., we refer to [16] or [21], whereas for the notion of *NP*-completeness and general background about algorithmic complexity we refer to [15] or [50].

F.1.3 Identifying codes

Identifying codes were introduced in [65] in 1998 with the original motivation of fault detection in multiprocessor systems. If G is a graph, an *identifying code* is a subset $\mathscr{C} \subseteq V(G)$ such that the family

$$\{N_G[v] : \ v \in \mathscr{C}\}$$

is an identifying system of $V(G)$. The elements of $\mathcal{C}$ are usually called *codewords*.

Of course, such a system will exist if and only if the family of all closed neighbourhoods $\{N_G[v] : v \in V(G)\}$ is itself identifying, which means in this case that distinct vertices must have distinct closed neighbourhoods ; a graph with this property is called *twin-free* or *identifiable*.

As aforementioned, in the original motivation the graph models a finite network of processors, and codewords correspond to processors equipped with a monitor able to detect a faulty processor in the closed neighbourhood of its location. Then, if there is at most one fault in the network and if every monitor sends a one-bit message referring to whether it detects a fault or not, we will be able to tell if there is a faulty processor in the graph, and locate it. See the graph G_1 on Figure I for an example ; one can check that a minimum identifying code in this graph has five codewords. Another example is the graph G_2, depicted on Figure II, which is a star on 15 vertices. One can check that the minimum size of an identifying code in G_2 is 14.

F.1.4 Watching systems

The graph G_1 (Figure I) is slightly pathological, because requiring five codewords to monitor six vertices is very much to ask (in fact, $n-1$ codewords is the maximum that can be required for a graph on n vertices, see [33] or [54]). The reason why we need so many codewords is that the closed neighbourhoods of two distinct vertices only differ by at most two vertices (the same phenomenon is also true for the leaves in G_2 on Figure II), and in this context a codeword has no choice but to check its whole closed neighbourhood, so that two distinct codewords check almost the same sets of vertices.

There are problems in which this situation is close to reality. For instance, consider a smoke detector : it has no choice but to detect smoke, regardless of the direction where it came from. So an identifying code is a good model for a fire-monitoring system in a building.

On the other hand, for instance in fault detection in multiprocessor systems, it seems plausible that we could easily assign a smaller control area to every detector by simply *not connecting it* to some adjacent vertices. Let us use the term *watcher* instead of codeword for this generalization.

First, let us define it informally with two examples. Assume that an edge between two vertices a and b denotes the possibility for a watcher in a to watch out what happens in b, but that we can choose not to use this possibility : thus we can assign to a watcher located at a vertex v a *watching zone*, which will be any subset of $N_G[v]$.

Let us try this on G_1 and check out Figure III : we only need three watchers with this protocol (the locations of the watchers 1, 2 and 3 are written down in squares, whereas the label of each vertex, i.e., the set of watchers watching it, is written down in italics nearby, so that the watching zone of each watcher can be retrieved), when five codewords were needed previously. All we have to check is that the labels of all vertices are nonempty and different, and that each watcher only watches vertices in the closed neighbourhood of its location.

What we can also do is to place several watchers at the same location, with distinct watching zones. For instance consider G_2 (see Figure IV) : only four watchers are needed whereas 14 codewords were necessary. This can be thought of as a single detector in the centre of the star, but needing four bits instead of one to send information, since it has 15 different vertices to watch. Thus watchers also enable us to model a monitoring system where monitors could simply tell where they detect a fault, but where the cost of a monitor is proportional to the number of bits needed to send this information.

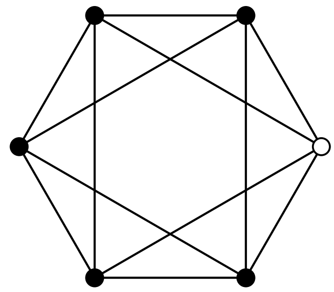


Figure I – *The graph G_1 and a minimum identifying code, of size five. Codewords are in black.*

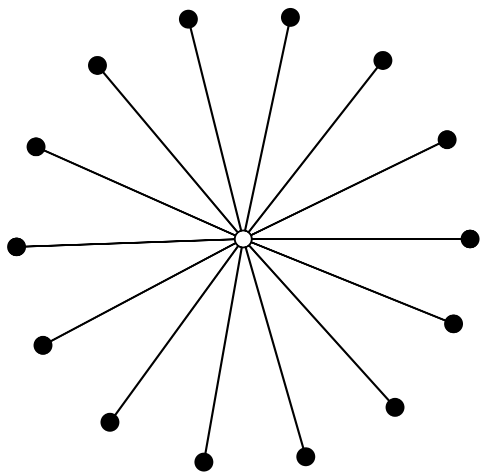


Figure II – *The graph G_2 and a minimum identifying code, of size 14. Codewords are in black.*

Let us define this formally :

Définition F.1. *A watching system in a graph $G = (V(G), E(G))$ is a finite set*

$$\mathcal{W} = \{w_1, w_2, \ldots, w_k\}$$

where each w_i is a couple $w_i = (v_i, Z_i)$, where v_i is a vertex and $Z_i \subseteq N_G[v_i]$, such that $\{Z_1, \ldots, Z_k\}$ is an identifying system.

Note that any graph G admits the trivial watching system $\{(v, \{v\}) : v \in V(G)\}$.

If $\mathcal{W}$ is a watching system in G and $w = (v, Z) \in \mathcal{W}$ is a watcher, we will say that v is the *location* of w, or that w is *located* at v. The set Z is the *watching zone*, or *watching*

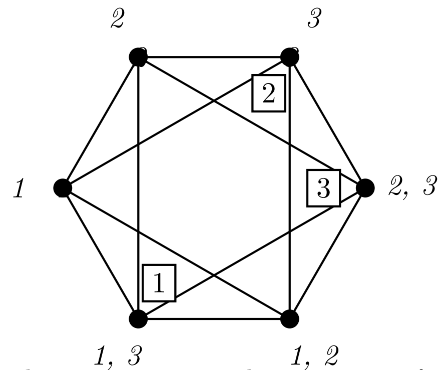


Figure III – *The graph G_1 with a minimum watching system, of size three. Watchers' locations are written down inside squares and labels nearby vertices, in italics.*

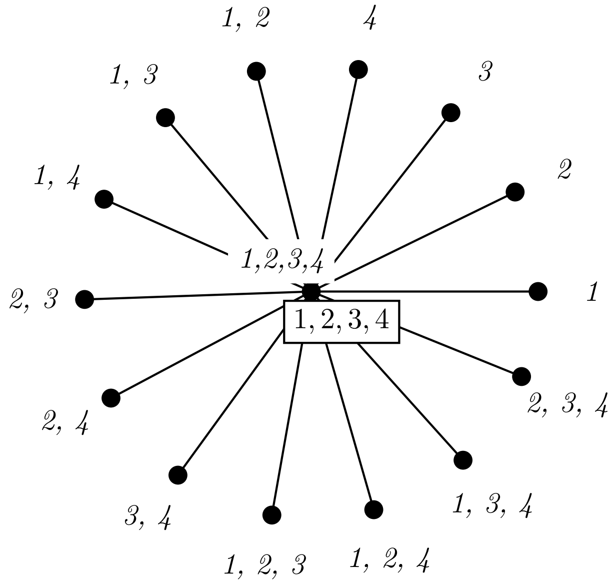


Figure IV – *A minimum watching system in G_2, of size four.*

area, of w, and if $x \in Z$ we say that w *covers* x, or that x is *covered by* w. We say that w *separates* the vertices x and y (or x from y) if w covers x and does not cover y, or the other way round. Therefore, $\mathcal{W}$ is a watching system of G if every vertex is covered by at least one watcher in $\mathcal{W}$ and any two distinct vertices are separated by at least one watcher in $\mathcal{W}$. Let us define the $\mathcal{W}$-*label*, or $\mathcal{W}$-*identifying set*, or simply *label*, of a vertex v as the set $L_{\mathcal{W}}(v)$ of watchers covering v. We will say that a vertex v is *identified* by $\mathcal{W}$ if its label $L_{\mathcal{W}}(v)$ is nonempty (v is covered by one watcher at least) and there is no other vertex in G with the same label. Thus another way to express the fact that $\mathcal{W}$ is a watching system is to say that all vertices in G are identified by $\mathcal{W}$.

F.2 First properties of watching systems

Let us recall that a *dominating set* in G is a subset Γ of $V(G)$ such that every vertex not in Γ is adjacent to at least one element in Γ. Let respectively $w(G)$, $\gamma(G)$ and $i(G)$ denote the minimum sizes of a watching system, of a dominating set and, when it exists, of an identifying code in G. These parameters will be called *watching number*, *domination* *number*, and *identifying number*, respectively.

If we have k questions to be answered by yes or no, there are $2^k - 1$ possibilities to answer all these questions without answering always by the negative, so we get a trivial lower bound for the size of a watching system. It is known that this bound also holds for identifying codes (see [65]). Noticing that an identifying code, when it exists, defines a watching system in an obvious way, we have the following relationship involving $|V(G)|$ and the watching and identifying numbers :

Theorem F.1. *For any graph G, we have :*

$$\lceil \log_2(|V(G)| + 1) \rceil \leq w(G).$$

For any twin-free graph G, we have :

$$w(G) \leq i(G).$$

We now compare the watching and domination numbers of a graph, with the following result, where $\Delta(G)$ denotes the maximum degree of G :

Theorem F.2. *For any graph G, we have :*

$$\gamma(G) \leq w(G) \leq \gamma(G) \cdot \lceil \log_2(\Delta(G) + 2) \rceil .$$

Proof. If $\mathcal{W}$ is a watching system, then the set of the watchers' locations in $\mathcal{W}$ is a dominating set, so we have the left-hand inequality. On the other hand, if we have a dominating set $\Gamma \subseteq V(G)$ of size $\gamma(G)$, we can identify all vertices simply by locating enough watchers at every vertex of Γ. One just has to notice that in order to identify a vertex v and its (at most) $\Delta(G)$ neighbours, we need at most $p := \lceil \log_2(\Delta(G) + 2) \rceil$ watchers, since a set with p elements has at least $\Delta(G) + 1$ nonempty subsets. □

F.3 An upper bound for the watching number

Since it is known that $i(G) \leq |V(G)| - 1$ for any connected graph with at least three vertices (see [33], [54]), this upper bound also holds for $w(G)$ by Theorem F.1. In fact, we prove much better in Theorem F.6, the proof of which will use the following three lemmas :

Lemma F.3. *Let G be a graph and H be a partial graph of G, i.e., with $V(H) = V(G)$ and $E(H) \subseteq E(G)$. Then*

$$w(H) \geq w(G).$$

Proof. If $\mathcal{W}$ is a watching system for H, then the same $\mathcal{W}$ is a watching system for G, since two adjacent vertices in H are also adjacent in G. □

Note that this monotony property does not hold in general for identifying codes.

Lemma F.4. *Let T be a tree, x be a leaf of T, and y be the neighbour of x.*

(a) *There exists a minimum watching system for T with one watcher located at y.*

(b) *If y has degree 2, there exists a minimum watching system for T with one watcher located at z, the second neighbour of y.*

Proof. (a) A watching system must cover x, so there is a watcher w_1 located at x or y, with $x \in Z$. If $w_1 = (x, Z)$, then we can replace it by $w_2 = (y, Z)$, since $N_G[y] \supseteq N_G[x]$.

(b) If $y \notin Z$, then one other watcher must cover y, and if $y \in Z$, then one must separate x and y, since $x \in Z$. In both cases, the task can be done by a watcher located at z. □

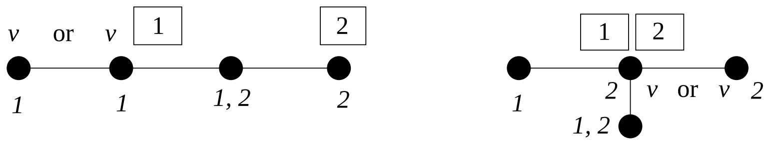


Figure V – *Trees with four vertices.*

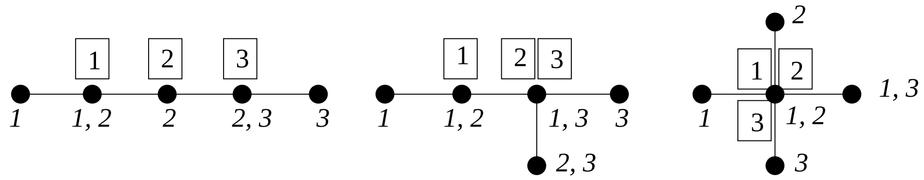


Figure VI – *The case n=5 in Theorem F.6.*

Lemma F.5. *Let T be a tree with four vertices, and let v be a vertex of T ; there exists a set $\mathcal{W}$ of two watchers such that*

- *the vertices in $V(T)\setminus\{v\}$ are covered and pairwise separated by $\mathcal{W}$ — in this case, we shall say, with a slight abuse of notation, that $\mathcal{W}$ is a watching system of $V(T)\setminus\{v\}$;*
- *the vertex v is covered by at least one watcher.*

Proof. On Figure V, we give all possibilities : the two trees with four vertices, and for each of them, the two possible locations for v (v is a leaf, or v is not a leaf). □

Theorem F.6. *Let G be a connected graph of order n, i.e., with n vertices.*

- *If $n = 1$, $w(G) = 1$.*
- *If $n = 2$ or $n = 3$, $w(G) = 2$.*
- *If $n = 4$ or $n = 5$, $w(G) = 3$.*
- *If $n \notin \{1, 2, 4\}$, $w(G) \leq \frac{2n}{3}$.*

Proof. For $n = 1$, $n = 2$, or $n = 3$, the result is direct. For $n = 4$, it is necessary to have at least $\lceil \log_2(5) \rceil = 3$ watchers and it is easy to verify that this is sufficient. For $n = 5$, all possibilities are given by Figure VI and we can see that we always have $w(G) = 3$.

We proceed by induction on n. We assume that $n \geq 6$ and that the theorem is true for any connected graph of order less than n.

Let G be a connected graph of order n. Let T be a spanning tree of G ; we will prove that $w(T) \leq \frac{2n}{3}$ and then the theorem will result from Lemma F.3. We denote by D the diameter of T and we consider a path $v_0, v_1, v_2, \ldots, v_{D-1}, v_D$ of T, with length D.

We distinguish between four cases, according to some conditions on the degrees of v_{D-1} and v_{D-2}.

• *First case : the degree of v_{D-1} is equal to 3*
The vertex v_{D-1} is adjacent to a vertex x other than v_{D-2} and v_D ; because D is the diameter, clearly x and v_D are leaves of T (see Figure VII). We consider the tree obtained by removing x, v_{D-1} and v_D from T ; this new tree T' has order $n - 3$.

If $n \geq 8$ or if $n = 6$, we consider a minimum watching system $\mathcal{W}$ for T' ; if $n = 7$, then T' is of order 4, and, using Lemma F.5, we choose a set $\mathcal{W}$ of two watchers which is a watching system for $V(T') \setminus \{v_{D-2}\}$ *and* covers the vertex v_{D-2}.

Then for T, in both cases, we add to $\mathcal{W}$ two watchers $w_1 = (v_{D-1}, \{v_{D-2}, v_{D-1}, v_D\})$ and $w_2 = (v_{D-1}, \{v_{D-1}, x\})$. On Figure VII, we rename 1 and 2 these watchers. Then $\mathcal{W} \cup \{w_1, w_2\}$ is a watching system for T. So, $w(T) \leq |\mathcal{W}| + 2 \leq w(T') + 2$.

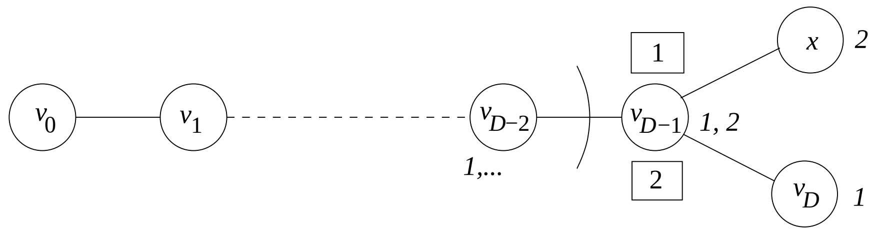


Figure VII – *First case of Theorem F.6 : the degree of v_{D-1} is equal to 3.*

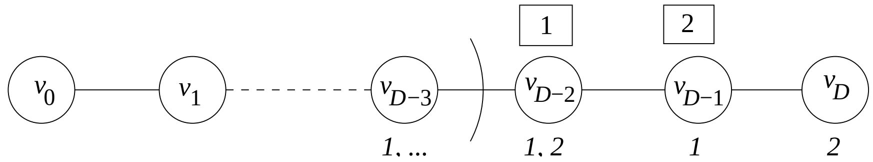


Figure VIII – *Second case of Theorem F.6 : the degrees of v_{D-1} and v_{D-2} are equal to 2.*

Now we use the induction hypothesis : if $n \geq 8$ or $n = 6$, then $w(T) \leq \frac{2}{3}(n-3)+2 = \frac{2n}{3}$; and if $n = 7$, then $w(T) \leq 2+2 = 4 < \frac{2}{3} \times 7$.

• *Second case : the degrees of v_{D-1} and v_{D-2} are equal to 2*
The neighbours of v_{D-1} are v_{D-2} and v_D, the neighbours of v_{D-2} are v_{D-3} and v_{D-1} (see Figure VIII). We consider the tree obtained by removing v_{D-2}, v_{D-1} and v_D from T ; this new tree T' has order $n-3$.

If $n \geq 8$ or if $n = 6$, we consider a minimum watching system $\mathcal{W}$ for T' ; if $n = 7$, T' is of order 4 ; again using Lemma F.5, we choose a set $\mathcal{W}$ of two watchers which is a watching system for $V(T') \setminus \{v_{D-3}\}$ and covers the vertex v_{D-3}. As in the first case, we add to $\mathcal{W}$ two watchers : $w_1 = (v_{D-2}, \{v_{D-3}, v_{D-2}, v_{D-1}\})$ and $w_2 = (v_{D-1}, \{v_{D-2}, v_D\})$, and obtain a watching system for T. So, $w(T) \leq |\mathcal{W}| + 2 \leq w(T') + 2$. The end of this case is the same as in the first case.

• *Third case : the degree of v_{D-1} is at least 4*
The vertex v_{D-1} is adjacent to at least two vertices other than v_{D-2} and v_D : let x and y be two neighbours of v_{D-1} distinct from v_{D-2} and v_D ; these two vertices are leaves of T (see Figure IX). We consider the tree T' obtained by removing x and y from T. By Lemma F.4, there exists a minimum watching system $\mathcal{W}$ of T' with a watcher w_1 located at v_{D-1}. For T, we take the set $\mathcal{W}$ and add the watcher $w_2 = (v_{D-1}, \{x, y\})$; we also add the vertex x to the watching zone of w_1. The set $\mathcal{W}$ being a watching system for T', the set $\mathcal{W} \cup \{w_2\}$ is a watching system for T. So, $w(T) \leq w(T') + 1$.

If $n \geq 7$, the order of T' is at least 5 and, using the induction hypothesis, $w(T) \leq \frac{2}{3}(n-2) + 1 < \frac{2n}{3}$.

If $n = 6$, then $n - 2 = 4$ and $w(T) \leq 3 + 1 = 4 = \frac{2}{3} \times 6$.

• *Fourth case : the degree of v_{D-1} is equal to 2 and the degree of v_{D-2} is at least 3*
The neighbours of v_{D-1} are v_{D-2} and v_D. The vertex v_{D-2} is adjacent to v_{D-3} and v_{D-1} but also to at least one other vertex x (see Figure X) ; if the degree of x is at least 3, using the fact that the diameter of T is equal to D, we can use the first or third case to conclude, with x playing the part of v_{D-1}.

So, we assume that the degree of x is 1 or 2 ; if its degree is 2, it has a neighbour y other than v_{D-2}.

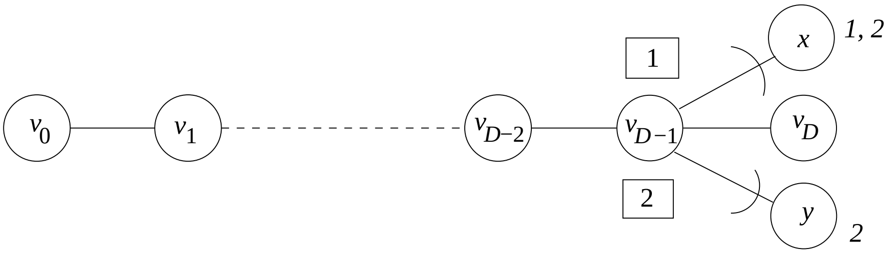


Figure IX – *Third case of Theorem F.6 : the degree of v_{D-1} is at least 4.*

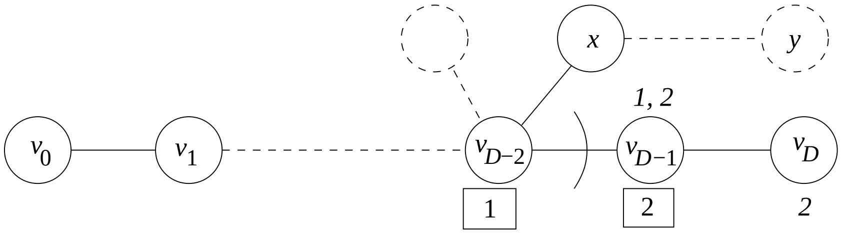


Figure X – *Fourth case of Theorem F.6 : the degree of v_{D-1} is equal to 2 and the degree of v_{D-2} is at least 3.*

We consider the tree T' of order $n-2$ obtained by removing v_{D-1} and v_D from T. By Lemma F.4, there exists a minimum watching system $\mathcal{W}$ of T' with a watcher w_1 located at v_{D-2}. For T, we take the set $\mathcal{W}$ and add the watcher $w_2 = (v_{D-1}, \{v_{D-1}, v_D\})$; we also add the vertex v_{D-1} to the watching zone of w_1. Then $\mathcal{W} \cup \{w_2\}$ is a watching system for T.

The end of this case is exactly the same as in the previous case. □

Moreover, we can almost characterize the graphs for which this bound is tight : in [8], we characterize the trees T with n vertices and $w(T) = \lfloor \frac{2n}{3} \rfloor$, then we characterize the graphs G with n vertices and $w(G) = \lfloor \frac{2n}{3} \rfloor$ in the cases $n = 3k$, $k \geq 1$, and $n = 3k+2$, $k \geq 1$; the case $n = 3k+1$ is more complex, and we are only able to state a conjecture for $k \geq 6$.

F.4 Watching systems in paths and cycles

Let us call a watching system $\mathcal{W}$ *compressed* if for every vertex $v \in V(G)$ and for every set A such that $\emptyset \subsetneq A \subsetneq L_{\mathcal{W}}(v)$, there is $v' \in V(G)$ such that $A = L_{\mathcal{W}}(v')$. See for example Figures III and IV.

If a watching system $\mathcal{W}$ is not compressed, then we can find v and A satisfying $\emptyset \subsetneq A \subsetneq L_{\mathcal{W}}(v)$ such that A is not the label of any vertex in G. Then if for every watcher (x, Z) in $L_{\mathcal{W}}(v) \setminus A$ we redefine this watcher by $(x, Z \setminus \{v\})$, we obtain another watching system of G where the labels of all vertices are the same as before, except for v that now has label A. Clearly, if we do this repeatedly we get a compressed watching system of G with the same size as $\mathcal{W}$, and thus we can always require a watching system to be compressed.

The following lemma is easy but will prove useful :

Lemma F.7. *Let G be a graph and $\mathcal{W}$ be a compressed watching system in G. Then for all $v \in V(G)$, we have :*

$$2^{|L_{\mathcal{W}}(v)|} - 1 \leq |B_G(v, 2)| .$$

Proof. Since $\mathcal{W}$ is compressed, all the $2^{|L_\mathcal{W}(v)|} - 1$ nonempty labels that can be formed using the watchers in $L_\mathcal{W}(v)$ must be attributed to vertices in G. The watchers in $L_\mathcal{W}(v)$ having their locations in $N_G[v]$, these labels can be attributed only inside $B_G(v, 2)$. □

The path P_n on n vertices is the graph whose vertex set is $\{1, 2, \dots, n\}$ and whose edge set is $\{\{i, i+1\} : 1 \leq i \leq n-1\}$. We prove :

Theorem F.8. *For all $n \geq 1$, we have :*

$$w(P_n) = \left\lceil \frac{n+1}{2} \right\rceil.$$

Proof. First let us prove that $w(P_n) \geq \frac{n+1}{2}$. Let $\mathcal{W}$ be a minimum compressed watching system of P_n and let i be such that $1 \leq i \leq n$. By Lemma F.7, since $|B_{P_n}(i, 2)| \leq 5$, we deduce that $|L_\mathcal{W}(i)| \leq 2$. Let us show that the vertices having a label of size 2 can be assumed to be nonadjacent.

First, assume that two adjacent vertices i and $i+1$ have respective labels ab and cd where a, b, c and d are distinct watchers. Since $\mathcal{W}$ is compressed, the four vertices around i and $i+1$ must be labeled by a, b, c and d (and thus we must also have $i > 2$ and $i+1 < n-2$). Without adding watchers, we can change $\mathcal{W}$ into a new compressed watching system where, without loss of generality, the labels from $i-2$ to $i+3$ are $a{-}ab{-}b{-}c{-}cd{-}d$.

Now assume that the labels of i and $i+1$ are ab and ac ; then the vertices with labels a, b and c must be in $i-2$, $i-1$, $i+2$ or $i+3$. If, for instance, the labels in this order are $b - a - ab - ac - c$, then we can replace them by $b - ab - a - ac - c$. It is not difficult to see that in all cases, we can get a watching system with the same size as $\mathcal{W}$, where the vertices with labels of size 2 are nonadjacent.

Let us also note that we can assume that the vertices 1 and n do not belong to this set : for instance, if the labels of 1, 2 and 3 are $ab{-}a{-}b$, we can replace them by $a{-}ab{-}b$, and a similar observation can be made for the vertices $n-2$, $n-1$, n.

Once $\mathcal{W}$ is modified, the set of vertices with a size-2 label is an independent set in the path $2, 3, \dots, n-1$ and thus has size at most $\left\lfloor \frac{n-1}{2} \right\rfloor$, and so the set of vertices with labels of size 1, whose cardinality is the same as $\mathcal{W}$, has size at least

$$n - \left\lfloor \frac{n-1}{2} \right\rfloor,$$

which is equal to $\left\lceil \frac{n+1}{2} \right\rceil$.

Constructions proving that $\left\lceil \frac{n+1}{2} \right\rceil$ is an upper bound are easy to find ; actually it is sufficient to use identifying codes (cf. [17]) : on the chains, watching systems are no better than identifying codes, except for $n = 2$, when no identifying code exists. □

The following result on cycles is obtained in a similar way. Let C_n denote the cycle of length n, with vertices $1, 2, \dots, n$, and edges $\{i, i+1\}$ for $i \in \{1, 2, \dots, n-1\}$, and $\{n, 1\}$.

Theorem F.9. *We have $w(C_4) = 3$, and for $n = 3$ and all $n \geq 5$:*

$$w(C_n) = \left\lceil \frac{n}{2} \right\rceil.$$

If we compare to identifying codes, we can see that the cycle of length three admits no identifying code and that $i(C_4) = i(C_5) = 3$; then $i(C_n) = \frac{n}{2}$ when n is even, $n \geq 6$ (see [17]), and $i(C_n) = \frac{n+3}{2}$ when n is odd, $n \geq 7$, see [42]. So $i(C_n) = w(C_n)$ when $n = 5$ or n is even, $n \geq 6$, and $i(C_n) = w(C_n) + 1$ when $n = 4$ or n is odd, $n \geq 7$.

F.5 Computational complexity

Let us recall what is a *vertex cover* in a graph G. An edge $e = xy \in E(G)$ is said to be *covered* by a vertex $v \in V(G)$ if v and e are incident, i.e., if $v = x$ or $v = y$. A *vertex cover* in G is a set of vertices $\mathscr{C} \subseteq V(G)$ such that every edge of G is covered by a at least one element $c \in \mathscr{C}$. Equivalently, $\mathscr{C}$ is a vertex cover if

$$\forall e = xy \in E(G), \quad x \in \mathscr{C} \text{ or } y \in \mathscr{C}.$$

It is well known that the problem of finding the minimum cardinality of a vertex cover in a given graph is *NP*-hard (see [64]) ; furthermore, it was proved in [49] that this problem remains *NP*-hard when restricted to the class of planar graphs whose maximum degree is at most 3, class which we denote by Π_3. For our proof we need to go a little further. In all graphs, a vertex of degree one is never an issue when we are looking for a vertex cover, since it is easy to prove the following lemma (see [7] for instance) :

Lemma F.10. *Let G be a graph and $xy \in E(G)$ be an edge such that the degree of the vertex x is 1, and let G' be the graph obtained by removing x, y and all their incident edges from G. Then the minimum cardinality of a vertex cover in G equals the minimum cardinality of a vertex cover in G' plus* 1.

Let Π'_3 be the class of all planar graphs where every vertex has degree 2 or 3. In addition to the aforementioned result from [49], Lemma F.10 proves that the following decision problem is *NP*-complete :

<u>MIN VERTEX COVER IN Π'_3</u>
- INSTANCE : A graph $G \in \Pi'_3$ and an integer k ;
- QUESTION : Is there a vertex cover for G with size at most k ?

We will use this *NP*-complete problem in order to study the computational complexity of the following decision problem :

<u>MIN WATCHING SYSTEM IN Π_3</u>
- INSTANCE : A planar graph G', with maximum degre at most 3, and an integer k' ;
- QUESTION : Is there a watching system for G' with size at most k' ?

We prove the following :

Theorem F.11. MIN WATCHING SYSTEM IN Π_3 *is NP-complete.*

Proof. Let us observe that MIN WATCHING SYSTEM IN Π_3 belongs to *NP*, since, given a watching system, it is polynomial with respect to the order of the graph to compute the labels of all vertices and check that they are nonempty and distinct. Now, using a polynomial reduction from the problem MIN VERTEX COVER IN Π'_3, we will show that our problem is *NP*-complete.

Consider a graph G and an integer k, an instance of the MIN VERTEX COVER IN Π'_3 problem. Denote respectively by n and m the number of vertices and edges of G. We construct a graph G' by replacing every edge xy of G by the structure S_{xy} depicted on Figure XI, consisting of 4 vertices (including x and y) and 3 edges. Thus G' has $n + 2m$ vertices and $3m$ edges and clearly the construction of G' from G can be done in polynomial time. Moreover, if $G \in \Pi'_3$, we clearly have $G' \in \Pi_3$. We set $k' = k + m$. The reduction will be complete if we prove that for all $k \geq 0$:

G admits a vertex cover of size at most k if and only if G' admits a watching system of size at most k'.

Consider an edge xy of G and the structure S_{xy} replacing xy in G', and let $V'_{xy} = \{a_{xy}, b_{xy}\}$.

Assume first that $\mathscr{C}$ is a vertex cover of G. We define a watching system $\mathcal{W}$ in G' as follows :

- for every vertex x of $V(G)$ such that $x \in \mathscr{C}$, we add the watcher $(x, N_{G'}[x])$ to $\mathcal{W}$;
- for every edge $e = xy$ of G, we add the watcher $(a_{xy}, N_{G'}[a_{xy}])$ to $\mathcal{W}$.

It is easy to see that $\mathcal{W}$ is a watching system in G'. Consider a vertex x in G ; since it has degree at least 2 in G, it is adjacent to at least two vertices y_1 and y_2 in G ; so the corresponding vertex x in G' is covered by two watchers located at a_{xy_1} and a_{xy_2}, belonging respectively to the structures S_{xy_1} and S_{xy_2}, and thus x is identified by $\mathcal{W}$. Also note that for every edge $e = xy$ of G, since either x or y belong to the vertex cover $\mathscr{C}$, there is a watcher in $\mathcal{W}$ that separates a_{xy} from b_{xy}. Thus G' admits a watching system with size $|\mathscr{C}| + m \leq k'$.

Conversely, assume that $\mathcal{W}$ is a watching system of G' of size at most k'. Consider an edge $xy \in E(G)$ and the watchers located in the structure S_{xy} of G'. Then :

- if no watcher is located at x nor y, there must be at least two watchers located in V'_{xy} ;
- if at least one watcher is located at x or y, we still need at least one watcher in V'_{xy}.

So if we denote by $\mathscr{C}$ the set of vertices $x \in V(G)$ such that $\mathcal{W}$ contains a watcher located at x, and by p the number of edges xy of G with $x \notin \mathscr{C}$ and $y \notin \mathscr{C}$, we have

$$|\mathscr{C}| \leq |\mathcal{W}| - 2p - (m - p) \leq k' - m - p \leq k - p.$$

Therefore if we add to $\mathscr{C}$ one vertex for every uncovered edge of G, we get a vertex cover of G of size at most k. □

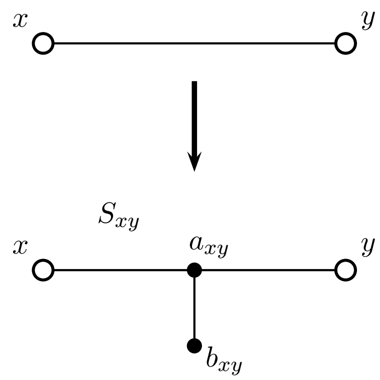


Figure XI – *The structure S_{xy} replacing every edge xy of G in the transformation.*

F.6 Distance-identification of sets of vertices

F.6.1 Definitions

Let us now turn to the problem of identifying several vertices within a certain distance, using a watching system. For $r \geq 1$ and $\ell \geq 1$, we define the notion of $(r, \leq \ell)$-watching systems which extends the notion of $(r, \leq \ell)$-identifying codes.

Define a r-watcher w in a graph G as in the case of a watcher $w = (v, Z)$ except for the watching zone Z that can now be any subset of the ball $B_G(v, r)$ centred at the location v of w and with radius r ; thus a 1-watcher is simply a watcher. We extend in an obvious way the notions of covering, label, separation, identification, ... to r-watchers. Any graph G admits the trivial $(r, \leq \ell)$-watching system $\{(v, \{v\}) : v \in V(G)\}$.

Let now $\mathcal{W}$ be a set of r-watchers in G. If $A \subset V(G)$, we define the $\mathcal{W}$-label of A as

$$L_{\mathcal{W}}(A) = \bigcup_{v \in A} L_{\mathcal{W}}(v),$$

and we say that $\mathcal{W}$ is a $(r, \leq \ell)$-*watching system* if all the labels of the subsets A of $V(G)$ with $1 \leq |A| \leq \ell$ are nonempty and distinct.

Note that a $(r, \leq \ell)$-watching system is a $(r', \leq \ell')$-watching system if $\ell' \leq \ell$ and $r' \geq r$.

Let $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ be a finite family of distinct nonempty subsets of a set X and $\ell \geq 1$. We say that $\mathcal{S}$ is a ℓ-*superimposed family* on X if, whenever we consider two distinct sets I, J included in $\{1, \ldots, k\}$ with $1 \leq |I| \leq \ell$ and $1 \leq |J| \leq \ell$, we have :

$$\bigcup_{i \in I} S_i \neq \bigcup_{j \in J} S_j.$$

This is the notion of ℓ-superimposed code in a set-system version. These codes were introduced in [66]. They are related to $(r, \leq \ell)$-identifying codes, as was observed in [65] and [48]. They are also related to watching systems since, with our definition, if $\mathcal{W}$ is a $(r, \leq \ell)$-watching system in a graph G, then the family of all $\mathcal{W}$-labels of the vertices of G is a ℓ-superimposed family on $\mathcal{W}$. Note that the family of singletons of X is always a ℓ-superimposed family of X for all $\ell \geq 1$, and so every graph G admits a $(r, \leq \ell)$-watching system for all $r \geq 1$ and $\ell \geq 1$, consisting of the watchers $(v, \{v\})$ for all $v \in V(G)$.

Observe that if $\ell \geq 2$ and $i \neq j$, then $S_i \subseteq S_j$ is impossible in a ℓ-superimposed family. From this follows that if $|L_{\mathcal{W}}(x)| = 1$ for a vertex x in the graph with watching system $\mathcal{W}$, then if $\ell \geq 2$ the watcher covering x must cover only x : we will call such a watcher an *hermit*. Without loss of generality, we can suppose that this watcher is $(x, \{x\})$, since its location does not matter.

F.6.2 The case of $(1, \leq 2)$-watching systems in paths and cycles

Let us start with the following lemma.

Lemma F.12. *For $1 \leq k \leq 4$, the only 2-superimposed family on a set with k elements with at least k subsets is the family of k singletons.*

Proof. The result is obvious if $1 \leq k \leq 3$, so we just check the case $k = 4$. Let S_1, S_2, S_3, S_4 be a 2-superimposed family on $\{1, 2, 3, 4\}$. If there is a singleton in the family, say $S_1 = \{1\}$, then we have $S_i \subset \{2, 3, 4\}$ for $i > 1$ and we use the case $k = 3$ to conclude.

If an element, say 1, is in at least three different sets, say S_1, S_2 and S_3, then by intersecting these sets with $\{2, 3, 4\}$ we get a 2-superimposed family of size 3 on $\{2, 3, 4\}$, so, using the case $k = 3$, we must have (up to permutations) $S_1 = \{1, 2\}$, $S_2 = \{1, 3\}$ and $S_3 = \{1, 4\}$. Then S_4 cannot contain 1, and the remaining possibilities for S_4 all lead to contradictions.

If all the elements are in at most two sets and there are no singletons, then by a simple counting argument we see that all the sets must be pairs, and so the family must be (up to permutations) $\{\{1, 2\}, \{1, 3\}, \{3, 4\}, \{2, 4\}\}$, which is not 2-superimposed. □

In other words, with k watchers, $1 \leq k \leq 4$, we can produce k valid labels, which will be singletons, and not more.

Theorem F.13. *For all $n \geq 1$, the minimum size of a $(1, \leq 2)$-watching system in the path P_n is at least $\frac{5}{6}n$.*

Proof. Let us again denote the vertices of the path P_n by $1, 2, \ldots, n$. First we check the result if $1 \leq n \leq 7$.

n	1	2	3	4	5	6	7
$\lceil 5n/6 \rceil$	1	2	3	4	5	5	6

If $1 \leq n \leq 4$, the result comes directly from Lemma F.12 : we need n watchers since the labels of the vertices form a 2-superimposed family of size n.

If $n = 5$ and we try with 4 watchers, then using Lemma F.12 on $\{1, 2, 3, 4\}$ and $\{2, 3, 4, 5\}$, we get a contradiction.

If $n = 6$, we need, as for $n = 5$, at least five watchers and the bound is satisfied.

If $n = 7$, consider the following argument : we need three watchers in $\{1, 2, 3, 4\}$ if we want to identify the sets $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$, and two other watchers in $\{5, 6, 7\}$ to identify $\{6\}$, $\{7\}$ and $\{6, 7\}$. So if only five watchers were to be used, then using Lemma F.12 twice, for $k = 3$ and for $k = 2$, five vertices would have a singleton for label, and no more labels would be available for other vertices. So we need at least six watchers and the bound is satisfied.

Assume now that $n \geq 8$ and that the result is true for smaller values of n. Consider a $(1, \leq 2)$-watching system $\mathcal{W}$ in P_n. We can make the following two assumptions :

- There are at least k watchers located in $\{1, 2, \ldots, k\}$ if $2 \leq k \leq 5$, and the same is true in $\{n-k+1, \ldots, n-1, n\}$. Indeed, if at most $k-1$ watchers were located in $\{1, \ldots, k\}$, by Lemma F.12 the labels of the vertices $1, 2, \ldots, k-1$ could only be singletons. In particular, no watcher in $\{1, 2, \ldots, k-1\}$ would cover vertices in $\{k, \ldots, n\}$ and vice versa ; so we could use induction on these two smaller paths, with $k-1$ vertices and $n-k+1$ vertices, to obtain the bound.
- There are at least five watchers located in $\{k, k+1, \ldots, k+5\}$ for every $1 \leq k \leq n-5$, by a similar argument : if we were to have at most four watchers to identify the vertices $k+1$, $k+2$, $k+3$ and $k+4$, then the labels of these vertices would be singletons and these watchers would not cover vertices in $\{1, \ldots, k\}$ nor in $\{k+5, \ldots, n\}$; so we could use induction with the paths $\{1, \ldots, k\}$ and $\{k+5, \ldots, n\}$.

With these assumptions we can easily prove the result : write $n = 6k + \rho$ with $0 \leq \rho \leq 5$, and cut P_n into k pieces with six vertices and one piece with ρ vertices at the end of the path.

If $\rho = 0$, since every piece with six vertices contains five watchers, the conclusion follows.

If $2 \leq \rho \leq 5$, the conclusion also follows since the last ρ vertices contain ρ watchers.

If $\rho = 1$, then we write $n = 3 + 6(k-1) + 4$ and need at least $3 + 5(k-1) + 4$ watchers. □

We offer a construction that matches asymptotically the bound (see Figure XII). It is a $(1, \leq 2)$-watching system of size $5k$ in a path of length $6k-1$, $k \geq 2$.

Observe also that no $(1, \leq 2)$-identifying code (and more generally, no $(1, \leq \ell)$-identifying code) exists in the path P_n, because, since $N_{P_n}[1] \subseteq N_{P_n}[2]$, the sets of vertices $\{2\}$ and $\{1, 2\}$ cannot be separated.

In the case of cycles, we have the following result.

Theorem F.14. *For all $n \geq 1$, the minimum size of a $(1, \leq 2)$-watching system in the cycle C_n is at least $\frac{5}{6}n$.*

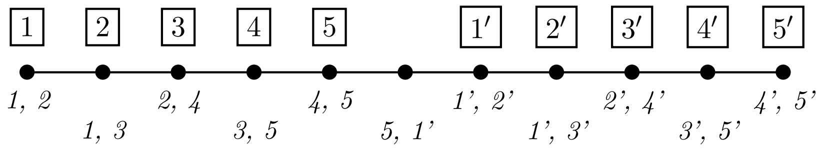


Figure XII – *An asymptotically optimal* $(1, \leq 2)$*-watching system in the path* P_n*.*

Proof. Mimicking the above proof for the path, we can see that the result is true for $1 \leq n \leq 6$, and that, when $n \geq 7$, every sequence of six consecutive vertices contains at least five watchers.

Consider now a watching system $\mathcal{W}$ of C_n, $n \geq 7$. We consider the n sequences B_i, $i \in \{1, 2, \ldots, n\}$, of six consecutive vertices, and let, for any watcher $w \in \mathcal{W}$, $\mathcal{F}_i(w) = 1$ if $w \in B_i$, $\mathcal{F}_i(w) = 0$ otherwise. Since any vertex belongs to exactly six sets B_i, and any set B_i contains at least five watchers, we have :

$$5n \;\leq\; \sum_{1 \leq i \leq n} \; \sum_{w \in \mathcal{W}} \mathcal{F}_i(w) \;=\; \sum_{w \in \mathcal{W}} \; \sum_{1 \leq i \leq n} \mathcal{F}_i(w) \;=\; 6|\mathcal{W}|,$$

and the claim follows. □

If we compare to $(1, \leq 2)$-identifying codes, we can see that, because $B_{C_n}(x, 1)$ and $B_{C_n}(x, 1) \cup B_{C_n}(x + 1, 1)$ differ by only one vertex, $x + 2$, this vertex, hence by symmetry all vertices, must belong to the code. Starting from $n = 7$, the only $(1, \leq 2)$-identifying code in the cycle C_n is $V(C_n)$.

F.6.3 The case of $(1, \leq \ell)$-watching systems in paths and cycles for $\ell \geq 3$

Like every graph, for all $n \geq 1$ the path P_n and the cycle C_n admit for all $\ell \geq 3$ a $(1, \leq \ell)$-watching system, which is the trivial watching system consisting of all the hermits. In the case of P_n and C_n, this is the best we can do :

Theorem F.15. *For all* $n \geq 1$ *and* $\ell \geq 3$*, the minimum size of a* $(1, \leq \ell)$*-watching system in the path* P_n *or the cycle* C_n *is* n*.*

Proof. Consider a $(1, \leq \ell)$-watching system $\mathcal{W}$ for P_n or C_n, where $\ell \geq 3$. Let $\mathcal{H} \subseteq \mathcal{W}$ be the set of hermits in $\mathcal{W}$, and let $V_{\mathcal{H}} \subset V(G)$ be the set of vertices covered by these hermits (as aforementioned, $V_{\mathcal{H}}$ can be taken, without loss of generality, as the set of the locations of the hermits). Now assume that there is a vertex x in $V(G) \setminus V_{\mathcal{H}}$; then we have $|L_{\mathcal{W}}(x)| > 1$. Suppose that $L_{\mathcal{W}}(x) = \{w_1, w_2\}$: then w_1 and w_2 are not hermits and so there exist a vertex $v_1 \neq x$ covered by w_1 and a vertex $v_2 \neq x$ covered by w_2 (though we may have $v_1 = v_2$). But in this case we have $L_{\mathcal{W}}(\{v_1, v_2\}) = L_{\mathcal{W}}(\{v_1, v_2, x\})$ (or $L_{\mathcal{W}}(\{v_1\}) = L_{\mathcal{W}}(\{v_1, x\})$ if $v_1 = v_2$), and so $\mathcal{W}$ cannot be a $(1, \leq \ell)$-watching system if $\ell \geq 3$.

Therefore, all vertices in $V(G) \setminus V_{\mathcal{H}}$ are covered by at least three watchers from $\mathcal{W} \setminus \mathcal{H}$; since a watcher can only cover at most three vertices, we clearly have $|\mathcal{W} \setminus \mathcal{H}| \geq |V(G) \setminus V_{\mathcal{H}}|$. Since we also have $|\mathcal{H}| = |V_{\mathcal{H}}|$, the result follows. □

Observe also that for $\ell \geq 3$, no $(1, \leq \ell)$-identifying code exists in the cycle C_n, because the sets of vertices $\{1, 3\}$ and $\{1, 2, 3\}$ (or more generally, $\{x, x + 2\}$ and $\{x, x + 1, x + 2\}$) cannot be separated.

Annexe G

Maximum Size of a Minimum Watching System and the Graphs Achieving the Bound

David Auger [1], **Irène Charon** [1], **Olivier Hudry** [1], **Antoine Lobstein** [2]

{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Abstract

Abstract Let $G = (V(G), E(G))$ be an undirected graph. A watcher w of G is a couple $w = (\ell(w),\ A(w))$, where $\ell(w)$ belongs to $V(G)$ and $A(w)$ is a set of vertices of G at distance 0 or 1 from $\ell(w)$. If a vertex v belongs to $A(w)$, we say that v is covered by w. Two vertices v_1 and v_2 in G are said to be separated by a set of watchers if the list of the watchers covering v_1 is different from that of v_2. We say that a set W of watchers is a watching system for G if every vertex v is covered by at least one $w \in W$, and any two vertices v_1, v_2 are separated by W. The minimum number of watchers necessary to watch G is denoted by $w(G)$. We give an upper bound on $w(G)$ for connected graphs of order n and characterize the trees achieving this bound, before studying the more complicated characterization of the connected graphs achieving this bound.

Keywords : Graph Theory, Watching Systems, Identifying Codes.

G.1 Introduction

Let $G = (V(G), E(G))$ be an undirected connected graph (the case of an unconnected graph can also be treated, by considering separately its connected components). A *watcher* w of G is a couple $w = (\ell(w),\ A(w))$, where $\ell(w)$ belongs to $V(G)$ and $A(w)$ is a set of

1. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

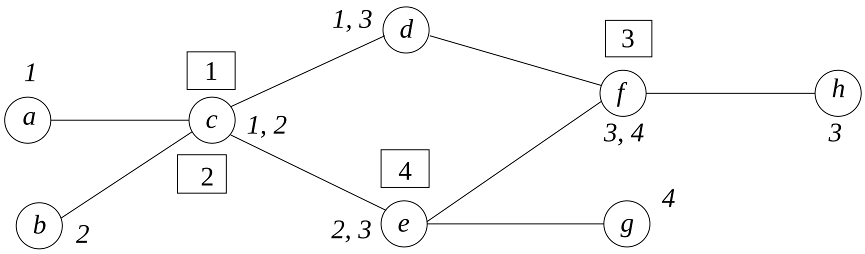


Figure I – *a graph G_0 watched by the watchers 1, 2, 3 and 4*

vertices of G at distance 0 or 1 from $\ell(w)$; in other words, $A(w)$ is a subset of $B(\ell(w))$, the ball of radius 1 centred at $\ell(w)$. We will say that w is *located* at $\ell(w)$ and that $A(w)$ is its *watching area* or *watching zone*. If a vertex v belongs to $A(w)$, we say that v is *covered* by w.

Two vertices v_1 and v_2 in G are said to be *separated* by a set of watchers if the list of the watchers covering v_1 is different from that of v_2.

We say that G is *watched* by a set W of watchers, or that W is a *watching system* for G if :

– for every v in $V(G)$, there exists $w \in W$ such that v is covered by w ;
– if v_1 and v_2 are two vertices of G, v_1 and v_2 are separated by W.

Note that several watchers can be located at a same vertex, and a watcher does not necessarily cover the vertex where it is located.

The minimum number of watchers necessary to watch a graph G is denoted by $w(G)$.

We will often represent watchers simply by integers, as for the graph G_0 represented in Figure I : the location of a watcher is written inside a rectangle ; for each vertex v of the graph, the list of watchers covering v is written in italics, so that the watching area of each watcher can be retrieved. In the example of Figure I, the watcher 1 is located at c and covers the vertices a, c and d, the watcher 2 is also located at c and covers the vertices b, c and e, the watcher 3 is located at f and covers the vertices d, e, f and h, and the watcher 4 is located at e and covers the vertices f and g. The graph G_0 is watched by these four watchers and it is easy to verify that $w(G_0) = 4$.

Let G be a graph of order n. If we have a set W of k watchers, the number of distinct non empty subsets of $\{\ell(w) : w \in W\}$ is equal to $2^k - 1$. Therefore, it is necessary to have $2^k - 1 \geq n$, and so we have the inequality :

$$w(G) \geq \lceil \log_2(n+1) \rceil. \tag{G.1}$$

Obviously, watching systems generalize *identifying codes* (see the seminal paper [65], and [73] for a large bibliography) : indeed, identifying codes are such that for any $w = (\ell(w), A(w)) \in W$, we have

$$A(w) = B(\ell(w)),$$

which means that, in this case, a watcher, or *codeword*, necessarily covers itself and all its neighbours.

See also [62], [83] for similar ideas.

Watching systems were introduced in [12], where motivations are exposed at large, basic properties are given, a complexity result is established, and the case of the paths is studied in detail, with comparison to identifying codes ; see also [13].

In Section G.2, we give an upper bound on $w(G)$ when G is a connected graph with n vertices. In Section G.3, we characterize the trees of order n which achieve this bound :

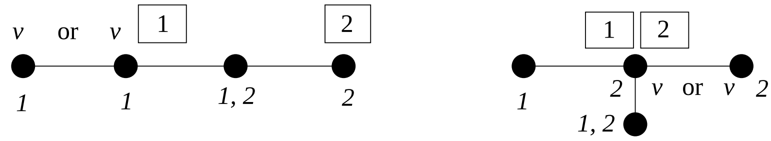


Figure II – *trees of order 4*

Theorems G.7, G.11 and G.12 are for the cases $n = 3k$, $n = 3k + 2$ and $n = 3k + 1$, respectively. This helps to study, in Section G.4, the characterization of *maximal* graphs reaching the bound, that is, graphs to which no edge can be added without decreasing the minimum number of necessary watchers : Theorems G.13 and G.14 give the answer for $n = 3k$ and $n = 3k + 2$ respectively, and Conjecture G.3 is stated for the case $n = 3k + 1$. This in turn gives results for all the connected graphs achieving the bound.

G.2 The maximum minimum number of watchers

The following three easy lemmata will prove efficient. We recall that $H = (V(H), E(H))$ is a partial graph of $G = (V(G), E(G))$ if $V(H) = V(G)$ and $E(H) \subseteq E(G)$.

Lemma G.1. *Let G be a graph and H be a partial graph of G. Then*

$$w(H) \geq w(G).$$

Proof. If H is watched by a set W of watchers, the same set W watches G, since two adjacent vertices in H are also adjacent in G. □

Note that this monotony property does not hold in general for identifying codes.

Lemma G.2. *Let T be a tree, x be a leaf of T, and y be the neighbour of x.*

(a) *There exists a minimum watching system for T with one watcher located at y.*

(b) *If y has degree 2, there exists a minimum watching system for T with one watcher located at z, the second neighbour of y.*

Proof. (a) A watching system must cover x, so there is a watcher w_1 located at x or y, with $x \in A(w_1)$. If $w_1 = (x, A(w_1))$, then we can replace it by $w_2 = (y, A(w_1))$, since $B_1(y) \supseteq B_1(x)$.

(b) If $y \notin A(w_1)$, then one other watcher must cover y, and if $y \in A(w_1)$, then one must separate x and y, since $x \in A(w_1)$. In both cases, the task can be done by a watcher located at z. □

Lemma G.3. *Let T be a tree of order 4 and let v be a vertex of T ; there exists a set W of two watchers such that*

- *the vertices in $V(T) \setminus \{v\}$ are covered and pairwise separated by W — in this case, we shall say, with a slight abuse of notation, that $V(T) \setminus \{v\}$ is watched by W ;*
- *the vertex v is covered by at least one watcher.*

Proof. On Figure II, we give all possibilities : the two trees of order 4, and for each of them, the two locations for v (v is a leaf, or v is not a leaf). □

We are now ready to give an upper bound for $w(G)$ with respect to n, the order of G. Note in contrast that the upper bound for identifying codes, when they exist, is $n - 1$, see [33], [54], and is reached, among other graphs, by the star.

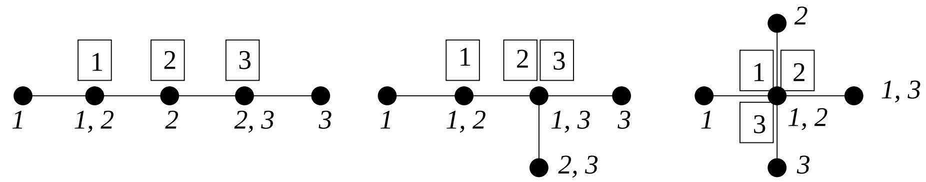


Figure III – *the case n=5 in Theorem G.4*

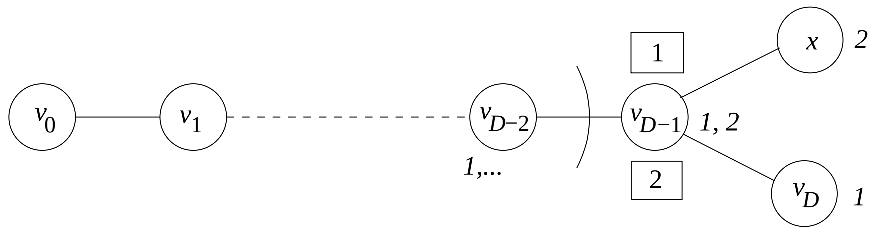


Figure IV – *first case of Theorem G.4 : the degree of v_{D-1} is equal to 3*

Theorem G.4. *Let G be a connected graph of order n.*
– *If $n = 1$, $w(G) = 1$.*
– *If $n = 2$ or $n = 3$, $w(G) = 2$.*
– *If $n = 4$ or $n = 5$, $w(G) = 3$.*
– *If $n \notin \{1, 2, 4\}$, $w(G) \leq \frac{2n}{3}$.*

The proof can be found in [12], [13], but we give it here, because the results of the four cases into which it is divided will be frequently used in the sequel.

Proof. For $n = 1$, $n = 2$, or $n = 3$, the result is direct. For $n = 4$, it is necessary to have at least $\lceil \log_2(5) \rceil = 3$ watchers and it is easy to verify that this is sufficient. For $n = 5$, all possibilities are given by Figure III and we can see that we always have $w(G) = 3$.

We proceed by induction on n. We assume that $n \geq 6$ and that the theorem is true for any connected graph of order less than n.

Let G be a connected graph of order n. Let T be a spanning tree of G ; we will prove that $w(T) \leq \frac{2n}{3}$ and then the theorem will result from Lemma G.1. We denote by D the diameter of T and we consider a path $v_0, v_1, v_2, \ldots, v_{D-1}, v_D$ of T, with length D.

We distinguish between four cases, according to some conditions on the degrees of v_{D-1} and v_{D-2}.

• *First case : the degree of v_{D-1} is equal to 3*
The vertex v_{D-1} is adjacent to a vertex x other than v_{D-2} and v_D ; because D is the diameter, clearly x and v_D are leaves of T (see Figure IV). We consider the tree obtained by removing x, v_{D-1} and v_D from T ; this new tree T' has order $n - 3$.

If $n \geq 8$ or if $n = 6$, we consider a minimum set W of watchers watching T' ; if $n = 7$, then T' is of order 4, and, using Lemma G.3, we choose a set W of two watchers to watch $V(T') \setminus \{v_{D-2}\}$ *and* cover the vertex v_{D-2}.

Then for T, in both cases, we add to W two watchers $w_1 = (v_{D-1}, \{v_{D-2}, v_{D-1}, v_D\})$ and $w_2 = (v_{D-1}, \{v_{D-1}, x\})$. On Figure IV, we rename 1 and 2 these watchers. Then T is watched by $W \cup \{w_1, w_2\}$. So, $w(T) \leq |W| + 2 \leq w(T') + 2$.

Now we use the induction hypothesis : if $n \geq 8$ or $n = 6$, then $w(T) \leq \frac{2}{3}(n-3)+2 = \frac{2n}{3}$; and if $n = 7$, then $w(T) \leq 2 + 2 = 4 < \frac{2}{3} \times 7$.

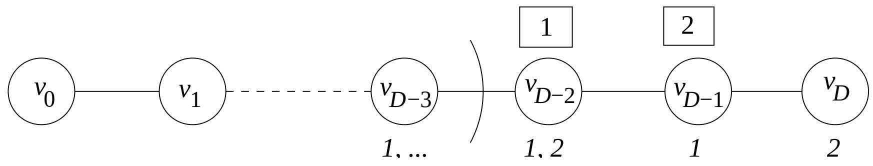


Figure V – *second case of Theorem G.4 : the degrees of v_{D-1} and v_{D-2} are equal to 2*

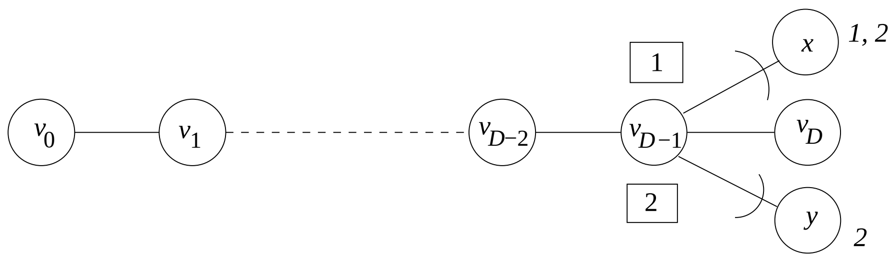


Figure VI – *third case of Theorem G.4 : the degree of v_{D-1} is at least 4*

• *Second case : the degrees of v_{D-1} and v_{D-2} are equal to 2*
The neighbours of v_{D-1} are v_{D-2} and v_D, the neighbours of v_{D-2} are v_{D-3} and v_{D-1} (see Figure V). We consider the tree obtained by removing v_{D-2}, v_{D-1} and v_D from T ; this new tree T' has order $n-3$.

If $n \geq 8$ or if $n = 6$, we consider a minimum set W of watchers watching T' ; if $n = 7$, T' is of order 4 ; again using Lemma G.3, we choose a set W of two watchers to watch $V(T') \setminus \{v_{D-3}\}$ and cover the vertex v_{D-3}. As in the first case, we add to W two watchers : $w_1 = (v_{D-2}, \{v_{D-3}, v_{D-2}, v_{D-1}\})$ and $w_2 = (v_{D-1}, \{v_{D-2}, v_D\})$, and T is watched. So, $w(T) \leq |W| + 2 \leq w(T') + 2$. The end of this case is the same as in the first case.

• *Third case : the degree of v_{D-1} is at least 4*
The vertex v_{D-1} is adjacent to at least two vertices other than v_{D-2} and v_D : let x and y be two neighbours of v_{D-1} distinct from v_{D-2} and v_D ; these two vertices are leaves of T (see Figure VI). We consider the tree T' obtained by removing x and y from T. By Lemma G.2, there exists a minimum set W of watchers watching T' with a watcher w_1 located at v_{D-1}. For T, we take the set W and we add the watcher $w_2 = (v_{D-1}, \{x, y\})$; we also add the vertex x to the watching area of w_1. The tree T' being watched by W, the tree T is watched by $W \cup \{w_2\}$. So, $w(T) \leq w(T') + 1$.

If $n \geq 7$, the order of T' is at least 5 and, using the induction hypothesis, $w(T) \leq \frac{2}{3}(n-2) + 1 < \frac{2n}{3}$.

If $n = 6$, then $n - 2 = 4$ and $w(T) \leq 3 + 1 = 4 = \frac{2}{3} \times 6$.

• *Fourth case : the degree of v_{D-1} is equal to 2 and the degree of v_{D-2} is at least 3*
The neighbours of v_{D-1} are v_{D-2} and v_D. The vertex v_{D-2} is adjacent to v_{D-3} and v_{D-1} but also to at least one other vertex x (see Figure VII) ; if the degree of x is at least 3, using the fact that the diameter of T is equal to D, we can use the first or third case to conclude, with x playing the part of v_{D-1}.

So, we assume that the degree of x is 1 or 2 ; if its degree is 2, it has a neighbour y other than v_{D-2}.

We consider the tree T' of order $n-2$ obtained by removing v_{D-1} and v_D from T. By

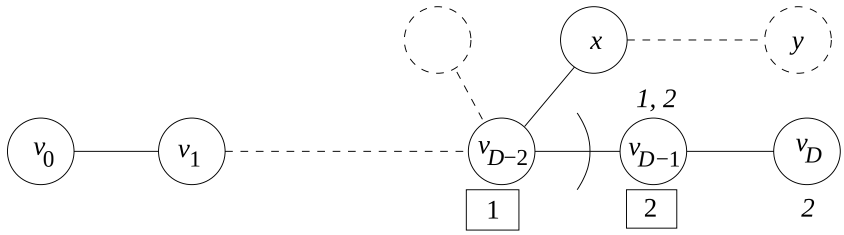


Figure VII – *fourth case of Theorem G.4 : the degree of v_{D-1} is equal to 2 and the degree of v_{D-2} is at least 3*

Lemma G.2, there exists a minimum set W of watchers watching T' with a watcher w_1 located at v_{D-2}. To watch T, we take the set W and add the watcher $w_2 = (v_{D-1}, \{v_{D-1}, v_D\})$; we also add the vertex v_{D-1} to the watching area of w_1. Then T is watched by $W \cup \{w_2\}$.

The end of this case is exactly the same as in the previous case. □

Remark G.5. *In the proof of Theorem G.4, we have constructed, according to the cases, a tree T' with order $n-3$ such that $w(T) \leq w(T') + 2$, or a tree T' with order $n-2$ such that $w(T) \leq w(T') + 1$.*

These two constructions, from T to T', will be used several times in the sequel, e.g., in the proof of Theorem G.7.

G.3 Trees T of order n achieving $w(T) = \lfloor \frac{2n}{3} \rfloor$

In this section, we characterize the trees T with n vertices and $w(T) = \lfloor \frac{2n}{3} \rfloor$. Our study is divided into three cases, $n = 3k$, $n = 3k+2$ and $n = 3k+1$.

We first define some particular trees, of order 1 to 5, that we name *gadgets*. For each gadget, we choose one or two particular vertex(ices) named *binding vertex(ices)*, through which the different gadgets will be exclusively connected between themselves ; a vertex which is not a binding vertex is said to be *ordinary*. In the sequel, we will sometimes denote a gadget of order i by gi, $1 \leq i \leq 5$, and use the abbreviation b. v. for binding vertex. The gadgets are depicted in Figure VIII ; we represent the binding vertices with squares and ordinary vertices with circles.

We will use the following easy lemma, whose proof we omit.

Lemma G.6. *Let T be a tree of order 3, and v and v' be two distinct vertices in T. It is possible to watch T with one watcher located at v and one watcher located at v'.*

As a consequence, if T' is a tree of order 4 and x is a leaf of T', there exists a set W of two watchers such that $T' \setminus \{x\}$ is watched by W and x is covered by W.

The following theorem characterizes the trees T with order $n = 3k$ and $w(T) = 2k$.

Theorem G.7. *Let T be a tree of order $n = 3k$ for $k \geq 1$. We have :*
$w(T) = 2k \Leftrightarrow T$ can be obtained by choosing k gadgets of order 3 and joining these gadgets by their binding vertices to obtain a tree.

The tree T_{15} in Figure IX is an example of a tree reaching this maximum.

Proof. Assume that a tree T of order $n = 3k$ is obtained by choosing k gadgets of order 3 and joining these gadgets by their b. v.'s to form a tree. It is clear that, to watch T, it is necessary to locate two watchers on each gadget. So T reaches the bound $2k$.

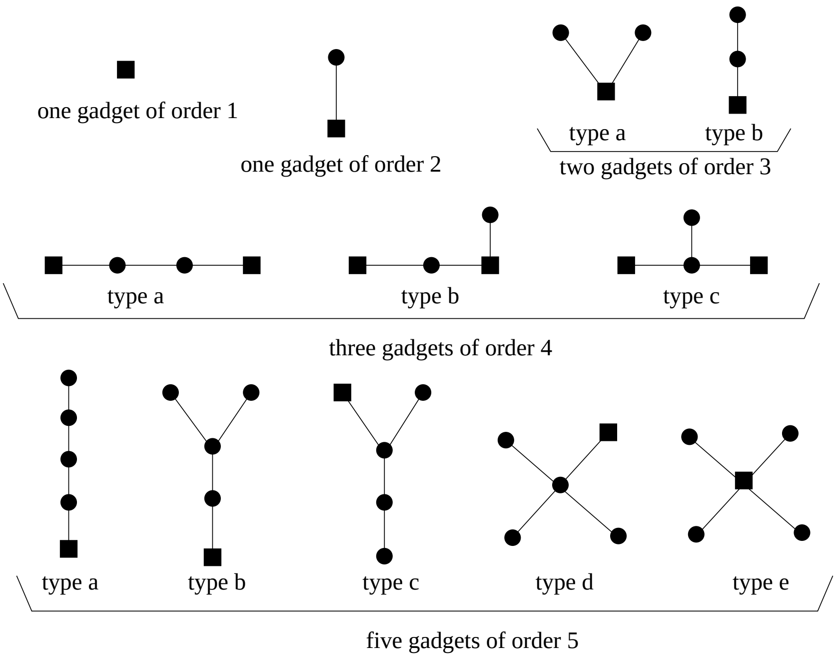


Figure VIII – *all the gadgets*

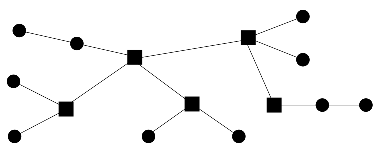


Figure IX – *the tree T_{15}*

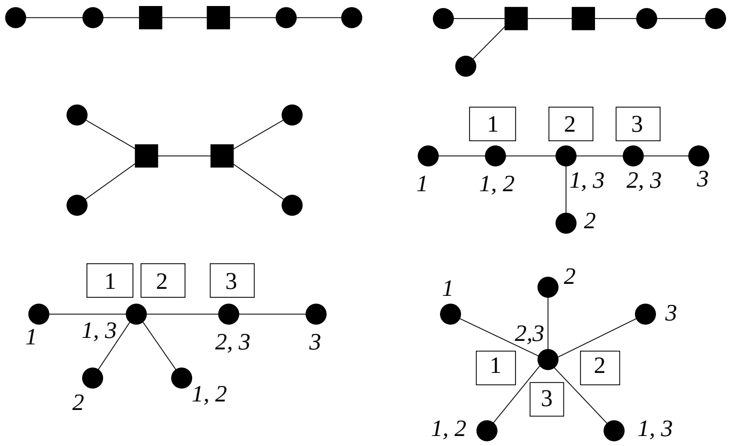


Figure X – *the trees of order 6 for the proof of Theorem G.7*

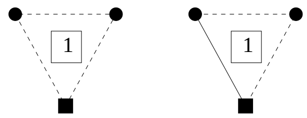


Figure XI – *two representations for a g3 of type a or b*

We will prove the converse by induction on k. For $k = 1$, it is immediate. We also examine the case $k = 2$, that is to say $n = 6$. We draw on Figure X the six different trees T on six vertices; when a tree is not of a type described in the right part of the equivalence, we explicitly give the watchers showing that $w(T) = 3$ and, in the other cases, we simply indicate the b. v.'s of the two gadgets involved.

We will sometimes represent a g3 of type a or b with a triangle, as on Figure XI : a dashed edge means that the edge may exist or not, with always exactly two edges in each g3. A watcher indicated inside the triangle means that this watcher is located at one of the three vertices of the triangle, at an appropriate vertex according to the case.

We assume now that $k \geq 3$ and that the theorem is true for $k' < k$. Let T be a tree of order $n = 3k$ with $w(T) = 2k$.

We consider again the proof of Theorem G.4 using a path $v_0, v_1, v_2, \ldots, v_{D-1}, v_D$ of length D, where D is the diameter of T. Here, the third and fourth cases are impossible, because they imply that $w(T) < \frac{2n}{3} = 2k$, unless $n = 6$, which has just been dealt with. In the first case of Theorem G.4, we rename by a, b, c and d respectively, the vertices v_{D-1}, v_D, x and v_{D-2}; in the second case, we rename by a, b, c and d respectively, the vertices v_{D-2}, v_{D-1}, v_D and v_{D-3}; in both cases, we remove the vertices a, b and c from T and obtain a tree T' of order $3(k-1)$; by Remark G.5, it appears that T' needs at least $w(T) - 2 = 2k - 2$ watchers and so $w(T') = 2(k-1)$ and we can apply the induction hypothesis to T' : the vertex d belongs to a g3, g.

Assume that d is not the binding vertex of g. The b. v. α of g is adjacent to the b. v. β of another g3 in T' (cf. Figure XII). By Lemma G.6, we can locate watchers w_4 and w_1 at a and β, so that d is covered by w_4 and α is covered by w_1; it is then possible to watch T with only one watcher located on the gadget g, as we can see on Figure XII, by choosing the appropriate vertex of g at which we locate the watcher denoted by 3. This leads to a contradiction on $w(T)$, and shows that d is the b. v. of g, in which case the result is immediately obtained, since $\{a, b, c\}$ can be seen as a g3, with its b. v. in a, connected

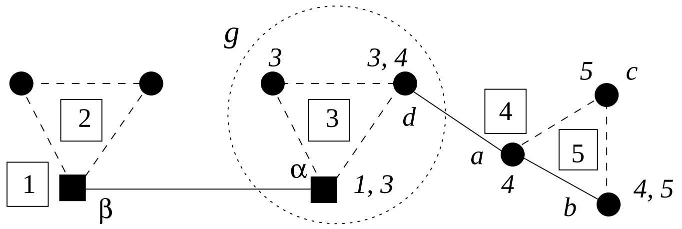


Figure XII – $2k-1$ *watchers are sufficient in T (end of proof of Theorem G.7)*

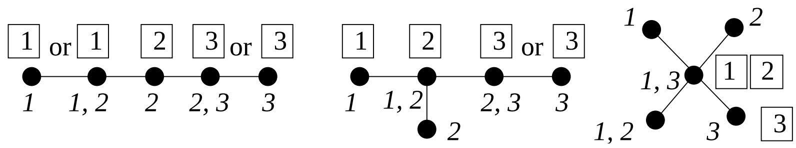


Figure XIII – *illustration for Lemma G.8*

to d. □

The following lemmata and definition will be used repeatedly in the sequel.

Lemma G.8. *Let T be a tree of order 5 and v be a vertex of T. It is possible to watch T with three watchers, one of the three watchers being located at v.*

As a consequence, if T' is a tree of order 6 and x is a leaf of T', there exists a set W of three watchers such that $T' \setminus \{x\}$ is watched by W and x is covered by W.

Proof. The result for T is straightforward, by examining all the different possibilities, as we can see on Figure XIII ; the consequence on T' is immediate. □

Lemma G.9. *Consider a g5 with binding vertex α and ordinary vertices v, x, y and z ; there exists a set W of two watchers such that*

- *$\{x, y, z\}$ is watched by W ;*
- *the vertex v is covered by W.*

Proof. If the g5 is of type a, b, c, or d, then the four vertices v, x, y, z form a tree, and by Lemma G.3, we are done. If the g5 is of type e, then it is also possible, with two watchers located at α, the centre of the star, to watch $\{x, y, z\}$ and cover v. □

Définition G.1. *Let $H = (V(H), E(H))$ be a connected graph and v be a vertex in $V(H)$; let H' be the graph obtained by removing the vertex v from H (H' is connected or not). We say that v is* free of charge*, or* free*, in H if there exists a minimum watching system for the graph H' which is also a watching system for H.*

Lemma G.10. *Let p be an integer verifying $p \geq 2$. Let F be a forest obtained by choosing p gadgets of order 3 or 5 and possibly, if desired, by adding edges between the binding vertices of the p gadgets. Let v be a new vertex, which is adjacent to at least one binding vertex and cannot be adjacent to ordinary vertices ; we assume that the graph obtained by adding v to F is a tree, T. Then, the vertex v is free in T.*

Proof. If v is adjacent to only one b. v., let α be this vertex ; since T is connected and $p \geq 2$, the vertex α is adjacent to another b. v., β. If v is adjacent to at least two b. v.'s among the p gadgets, let α and β be two such vertices. Figure XIV illustrates the lemma in detail in three cases :

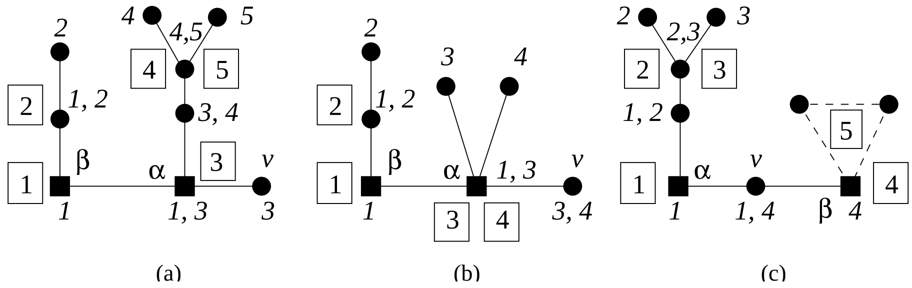


Figure XIV – *illustration for Lemma G.10*

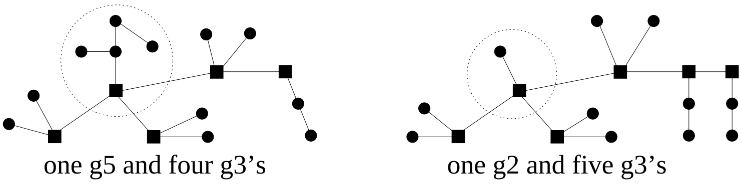


Figure XV – *the trees T_{17} and T'_{17}*

(a) v is linked to the b. v. α of a g5 and α is linked to the b. v. of a g3 ;
(b) v is linked to the b. v. α of a g3 and α is linked to the b. v. of a g3 ;
(c) v is linked to the b. v.'s of a g5 and of a g3.
The other cases, using repeatedly Lemmata G.6 and G.8, can be treated exactly in the same way. □

We are now ready to characterize the trees T with order $n = 3k + 2$ and $w(T) = 2k + 1$.

Theorem G.11. *Let T be a tree of order $n = 3k + 2$ for $k \geq 1$. We have :*
$w(T) = 2k + 1 \Leftrightarrow T$ can be obtained by choosing one gadget of order 2 and k gadgets of order 3, or one gadget of order 5 and $k - 1$ gadgets of order 3, and joining these gadgets by their binding vertices to obtain a tree.

The trees T_{17} and T'_{17} of Figure XV are examples of trees which achieve this maximum.

Proof. Assume that a tree T of order $n = 3k + 2$ is obtained by choosing one g2 and k g3's, or one g5 and $k-1$ g3's, and finally joining these gadgets by their binding vertices, in order to obtain a tree. It is necessary to locate one watcher on a g2, two watchers on a g3 and, because a g5 has four ordinary vertices, three watchers on a g5. So T achieves the bound $2k+1$: if there is a g2, we need one watcher for the g2 and $2k$ watchers for the k g3's, if there is a g5, three watchers for the g5 and $2k - 2$ watchers for the $k - 1$ g3's.

We will prove the converse by induction on k. For $k = 1$, $n = 5$ and the result is clear, see Figure III : T is a g5 (and in two out of three cases, it can also be seen as the connexion of a g2 and a g3). We assume now that $k \geq 2$ and that the theorem is true for $k' < k$. Let T be a tree of order $n = 3k + 2$ with $w(T) = 2k + 1$. We consider again the proof of Theorem G.4, using a path $v_0, v_1, v_2, \ldots, v_{D-1}, v_D$ of length D, where D is the diameter of T.

• *Part (a) : we assume that we are in the first or second case in the proof of Theorem G.4*
In the first case, we rename by a, b, c and d respectively, the vertices v_{D-1}, v_D, x and v_{D-2} ; in the second case, we rename by a, b, c and d respectively, the vertices v_{D-2}, v_{D-1},

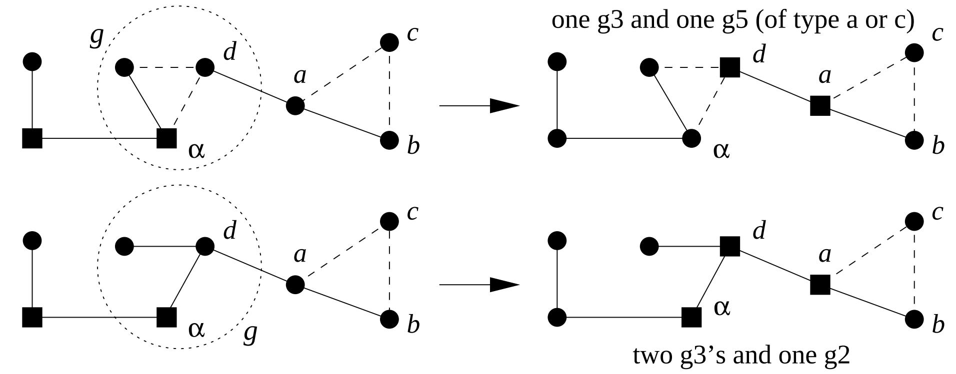


Figure XVI – *cases for $n = 8$ in part (a) of Theorem G.11, when g is of order 3*

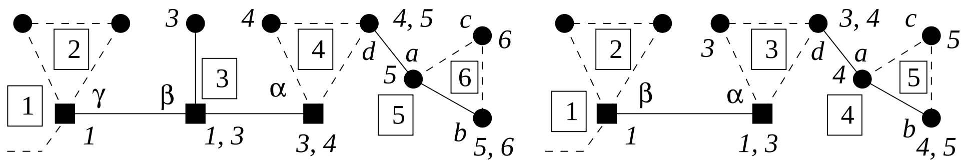


Figure XVII – *cases for $n \geq 11$ in part (a) of Theorem G.11, when g is of order 3*

v_D and v_{D-3} ; we remove the vertices a, b and c from T and obtain a tree T' of order $3(k-1)+2$; by Remark G.5, it appears that T' needs at least $w(T)-2 = 2k-1$ watchers and so $w(T') = 2(k-1)+1$ and we can apply the induction hypothesis to T' : T' is of one of the two types described in the right part of the equivalence, and the vertex d belongs to a gadget g, whose b. v. we denote by α. Assume first that $d \neq \alpha$.

• (i) If g is of order 2, the subtree induced by the vertices of g and the vertices a, b and c yields a g5 of type a or b, and the result is proved for T.

• (ii) Assume next that g is of order 3. If T is of order 8, the two possibilities are given by Figure XVI. If T is of order at least 11, there exists in T' a gadget g' connected to α, and we denote by β the b. v. of g'. If g' is a g2, which is itself connected to another g3 with binding vertex γ, then the left part of Figure XVII shows how to use only one watcher for g, which leads to a contradiction on $w(T)$; the same is true if g' is a g3, see the right part of Figure XVII, which actually is the same as Figure XII ; the case when g' is a g5 works similarly, using Lemma G.8. So the only case left is when g' is a g2 connected only to g, and Figure XVIII shows how to solve it.

• (iii) Finally, assume that g is of order 5. If T is of order 8, the reader will convince himself that locating d at all the different vertices, except at the b. v. α, of all the different types for a g5 leads to the six patterns given by Figure XIX. If T is of order 11, the b. v. α is adjacent to the b. v. β of a g3. When one examines the different possibilities, it appears that if T reaches the bound, it is of the wished shape : this is shown by Figure XX.

To close the case when g is of order 5, we study the case when T is of order at least 14 ; then the tree T'' obtained from T' by removing the four vertices of g other than α has order at least 7 and we can apply Lemma G.10 to it, which shows that the vertex α is free in T''. Using Lemma G.9, we can use two watchers on g to watch $V(g) \setminus \{\alpha, d\}$ and cover the vertex d. With one watcher at a covering d, we can separate d from all the other vertices : so, we can do with only two watchers on g, and T does not achieve the bound.

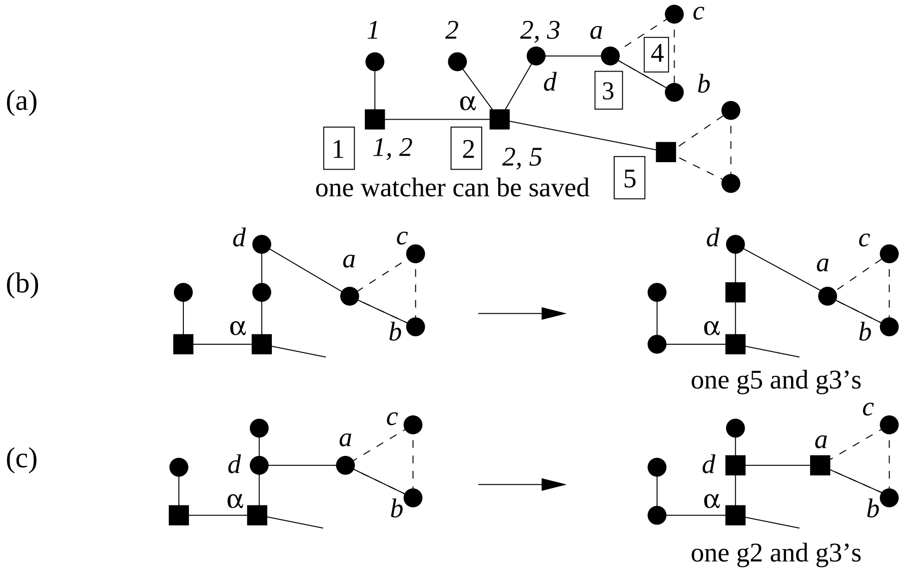


Figure XVIII – *more cases for $n \geq 11$ in part (a) of Theorem G.11, when g is of order 3*

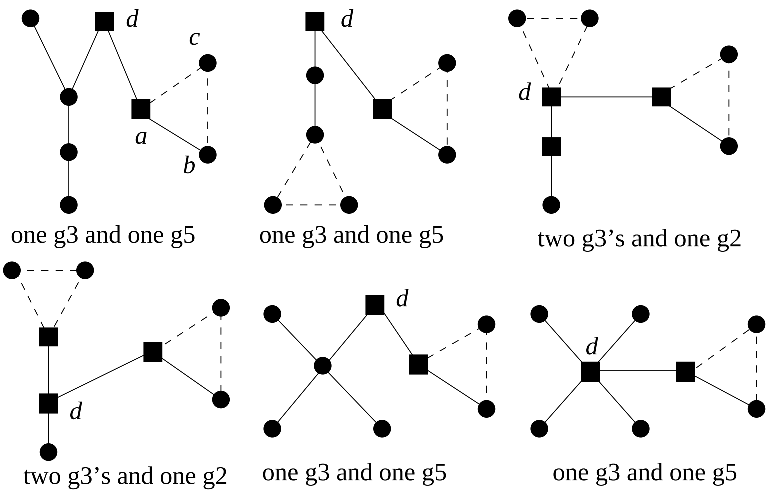


Figure XIX – *cases for $n = 8$ in part (a) of Theorem G.11, when g is of order 5*

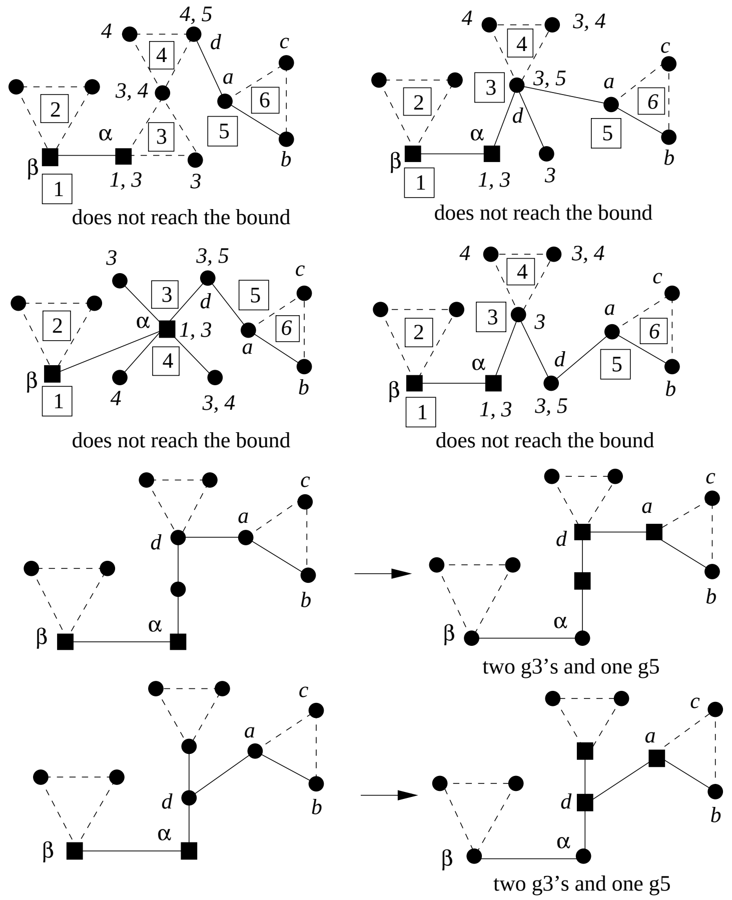


Figure XX − *cases for $n = 11$ in part (a) of Theorem G.11, when g is of order 5*

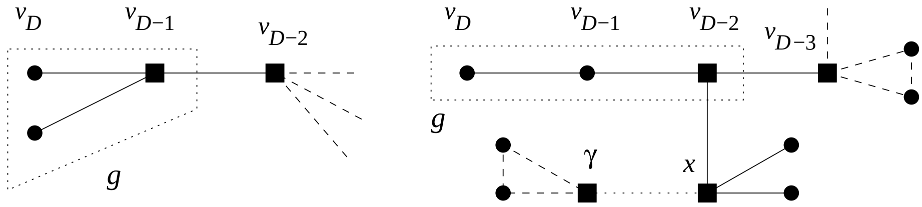


Figure XXI – *illustration for part (b) of Theorem G.11*

This shows that if $d \neq \alpha$, then either the tree does not achieve the bound, or it is of the desired form. On the other hand, if $d = \alpha$, then the result is immediately obtained. This ends part (a).

• *Part (b) : we assume that we are in the third or fourth case in the proof of Theorem G.4*
If we are in the third case, we remove the vertices x and y and if we are in the fourth case, we remove the vertices v_{D-1} and v_D ; we obtain a tree T' of order $3k$. By Remark G.5, we have $w(T') = 2k$ and Theorem G.7 may be used : T' can be obtained as a collection of g3's linked by some edges between their binding vertices. So, the vertex v_0 is a leaf of a g3, g ; now we reverse the longest path $v_0, v_1, \ldots, v_D$ in T. If g is of type a (see the left part of Figure XXI), then v_{D-1} is linked to only one b. v., v_{D-2}, and has degree 3, because D is the diameter of the tree, and we are brought back to the first case. And if g is of type b, then v_{D-1} has degree 2, and either v_{D-2} has degree 2 and we are in the second case, or v_{D-2} has degree at least 3 and we are in the fourth case, with at least one b. v. x linked to v_{D-2} and x of degree at least 2 (see the right part of Figure XXI) ; however, x cannot be linked to another b .v. γ, since this would increase the diameter of the tree, and for the same reason the g3 of x is of type a, so that necessarily x has degree 3. With x playing the part of v_{D-1}, we are again in the first case. In all cases, we can re-use the result obtained in part (a). □

The last case, $n = 3k + 1$ and $w(T) = 2k$, offers the greatest number of possibilities for the gadgets.

Theorem G.12. *Let T be a tree of order $n = 3k + 1$ for $k \geq 2$. We have :*
$w(T) = 2k \Leftrightarrow T$ *can be obtained by choosing*
– (i) *two gadgets of order 2 and $k - 1$ gadgets of order 3,*
– (ii) *or one gadget of order 2, one gadget of order 5 and $k - 2$ gadgets of order 3,*
– (iii) *or two gadgets of order 5 and $k - 3$ gadgets of order 3,*
– (iv) *or one gadget of order 1 and k gadgets of order 3,*
– (v) *or one gadget of order 4 and $k - 1$ gadgets of order 3,*
and joining these gadgets by their binding vertices to obtain a tree.

It may be that one of the two binding vertices of a g4 is not linked to any (binding) vertex. The trees T_{13}, T'_{13} and T''_{13} of Figure XXII are examples of trees achieving the bound $2k$ for $n = 3k + 1$ (with $k = 4$).

Proof. Assume that a tree T of order $n = 3k+1$ is obtained as specified in the right part of the above equivalence. It is necessary to locate one watcher on a g2, two watchers on a g3 and two on a g4 (because a g4 has two ordinary vertices), and three watchers on a g5. So T reaches the bound $2k$:
if we are in (i), $(2 \times 1) + ((k - 1) \times 2) = 2k$;
in (ii), $(1 \times 1) + (1 \times 3) + ((k - 2) \times 2) = 2k$;

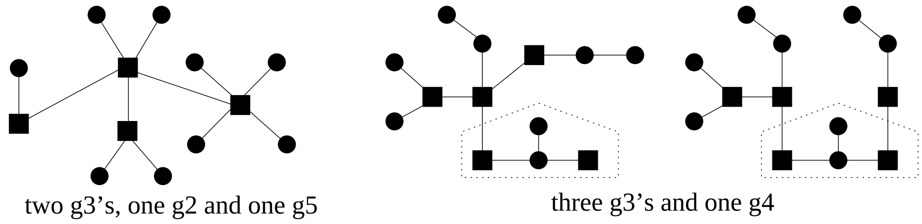


Figure XXII – *the trees T_{13}, T'_{13} and T''_{13}*

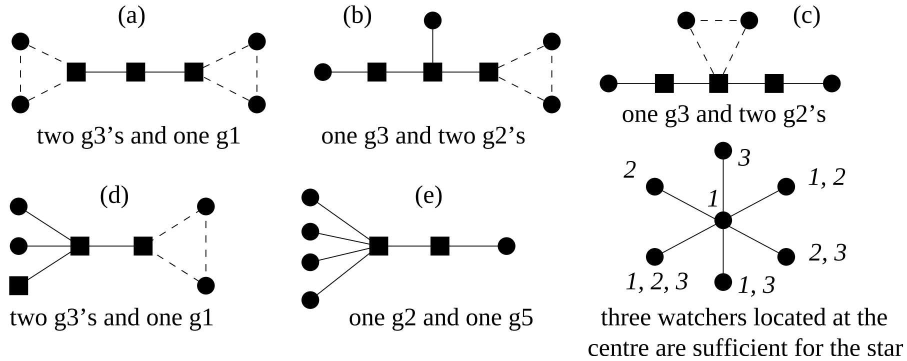


Figure XXIII – *all the possibilities for $n = 7$ in Theorem G.12*

in (iii), $(2 \times 3) + ((k-3) \times 2) = 2k$;
in (iv), $(1 \times 0) + (k \times 2) = 2k$;
in (v), $(1 \times 2) + ((k-1) \times 2) = 2k$.
We will prove the converse by induction on k. For $n = 7$, the different possibilities are examined on Figure XXIII. Now, we assume that $n \geq 10$.

We use the same scheme of proof as for Theorem G.11 : we assume that $k \geq 3$ and that the theorem is true for $k' < k$, we let T be a tree of order $n = 3k+1$ with $w(T) = 2k$, and we consider the proof of Theorem G.4, using a path $v_0, v_1, v_2, \ldots, v_{D-1}, v_D$ of length D, where D is the diameter of T.

• *Part (a) : we assume that we are in the first or second case in the proof of Theorem G.4*
If in T there is a g4, and if only one of its binding vertices is connected to another gadget, this g4 can be viewed as two g2's or as one g1 and one g3. So we can assume that, if there is a g4, i.e., if we are in case (v) of the theorem, then each of the two b. v.'s of the g4 is connected to at least one g3.

In the first case, we rename by a, b, c and d respectively, the vertices v_{D-1}, v_D, x and v_{D-2} ; in the second case, we rename by a, b, c and d respectively, the vertices v_{D-2}, v_{D-1}, v_D and v_{D-3}. In both cases, we remove the vertices a, b and c from T and obtain a tree T' with order $3(k-1)+1$; by Remark G.5, it appears that T' needs at least $w(T)-2 = 2k-2$ watchers and so $w(T') = 2(k-1)$: we can apply the induction hypothesis to T', which is of one of the five types described in the right part of the equivalence.

The vertex d belongs to a gadget g ; as before, if d is a binding vertex, we are done, so we assume from now on that d is ordinary, so that g is of order 2 or more, and we have four cases, according to the order of g.

• (1) If g is of order 2, the subtree induced by the vertices of g and the vertices a, b and c yields a g5 and the result is proved : indeed, if T' has two g2's and $k-2$ g3's

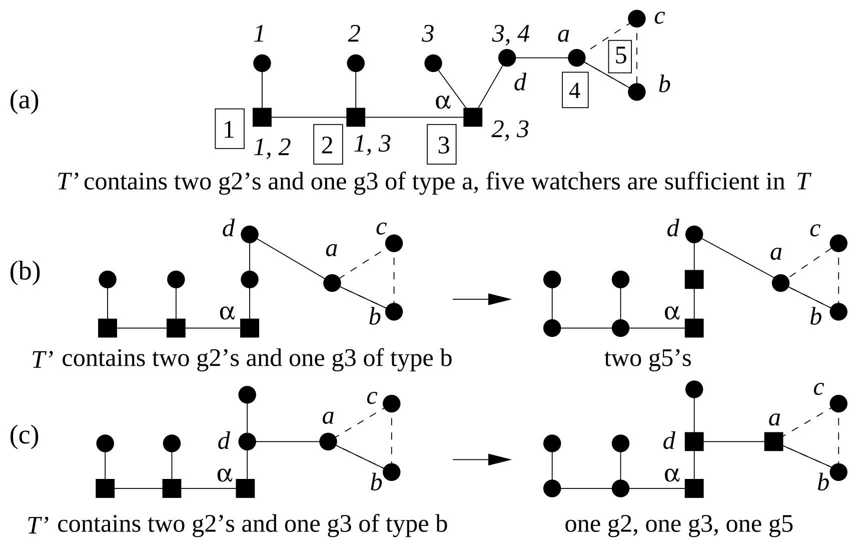


Figure XXIV – *cases for* $n = 10$ *in part (a) of Theorem G.12, when* g *is a g3*

(case (i)), or one g2, one g5 and $k-3$ g3's (case (ii)), then T can be obtained with one g2, one g5 and $k-2$ g3's (case (ii)), or two g5's and $k-3$ g3's (case (iii)), respectively.

• (2) Assume next that g is of order 3, with binding vertex α.

If $n = 10$, we consider for T' the cases in Figure XXIII where there is at least one g3, that is, the cases (a)–(d). The cases (a) and (d), where there is a g1, will be studied below, for general values of n. If, in the case (b), g is of type a, then, see Figure XXIV(a), five watchers are sufficient to watch T, whereas if g is of type b, then, according to the location of d in g, T consists of two g5's, or of one g2, one g3 and one g5, see Figure XXIV(b) and (c).

Similarly, if in the case (c) of Figure XXIII, g is of type a, then five watchers are sufficient, whereas if g is of type b, then T consists of one g2, one g3 and one g5, or of two g3's and two g2's — cf. Figure XXX(a) below.

We consider now the case when there is a g1, with vertex δ, in T', with $n \geq 10$: we are in case (iv) of Theorem G.12 and all the other gadgets in T' are g3's. If δ is linked neither to α nor to any neighbour of α, we are in the situation depicted by Figure XXV(a) and $2k-1$ watchers are sufficient for T : since γ or δ is linked to another g3, these two vertices can be separated by another watcher. So we can assume from now on that δ is linked either to α or to one of its neighbours. First, we assume that α is linked to at least one g3, cf. the left part of the tree in Figure XXV(a). Again, we can save one watcher, so that T does not achieve the bound $2k$, unless we are in one of the following three cases :

(i) δ is linked only to α, see Figure XXV(b). Then either δ is not covered by any watcher, or it is covered by the watcher 3 located at α, in which case no watcher separates e and δ. This gives the three possibilities detailed in Figure XXVI.

(ii) δ is linked to α and to exactly one other g3, which is not linked to any other g3, and the watcher 3 cannot be located at α, which means that g is of type b, see Figure XXV(c), where γ and δ are not separated. Then there are two possibilities, given in Figure XXVII.

(iii) δ is linked to a neighbour β of α, and neither δ nor β is linked to other b. v.'s, and g is of type b, see Figure XXV(d), where β and f are not separated. Figure XXVIII shows then that we still are in the conditions of Theorem G.12.

Now we can assume that α is not linked to any g3, which means that it is linked to δ,

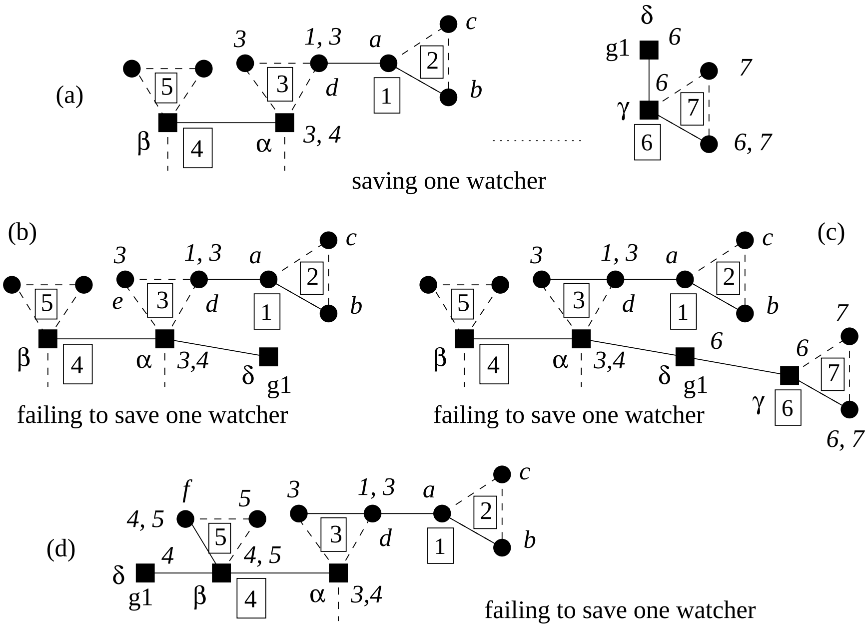


Figure XXV – *part (a) of Theorem G.12 : g is a g3 and there is a g1*

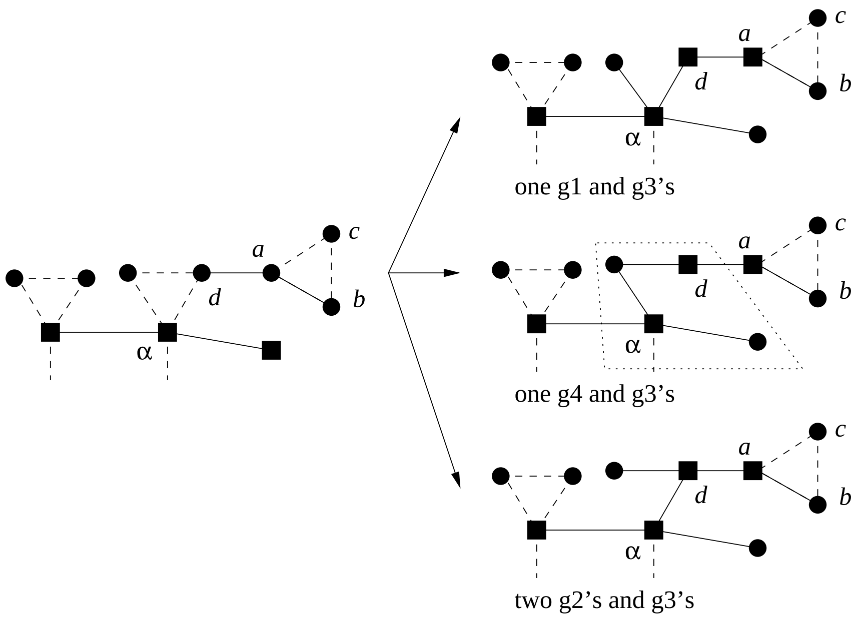


Figure XXVI – *cases with a g1 in part (a) of Theorem G.12, when g is a g3*

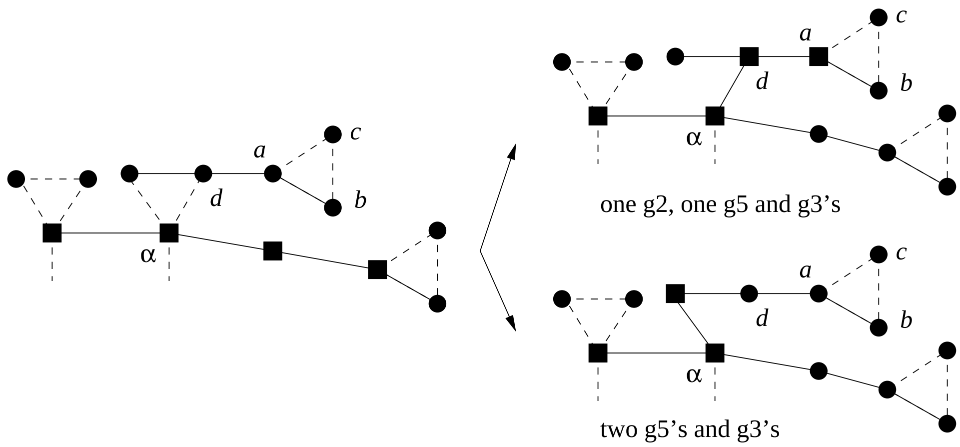


Figure XXVII – *more cases with a g1 in part (a) of Theorem G.12, when g is a g3*

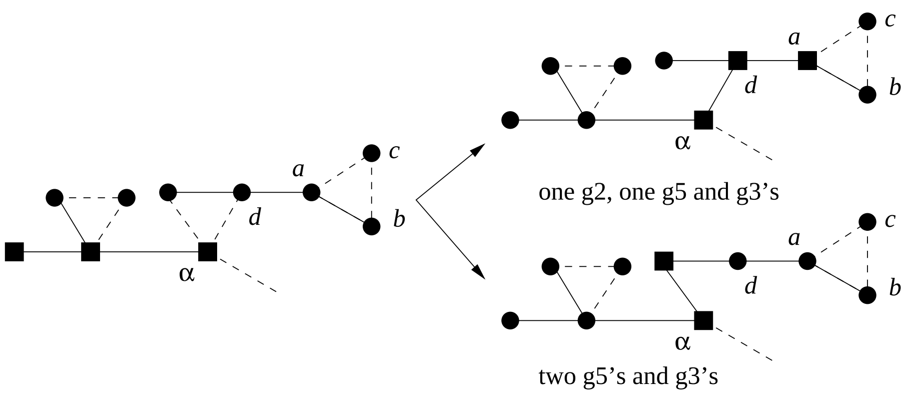


Figure XXVIII – *last cases with a g1 in part (a) of Theorem G.12, when g is a g3*

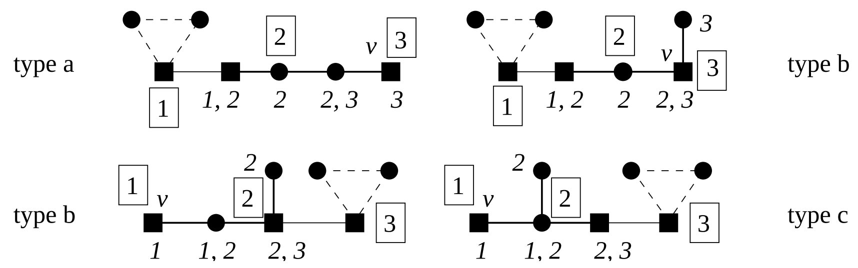


Figure XXIX – *choice of a watcher at a binding vertex of a g4*

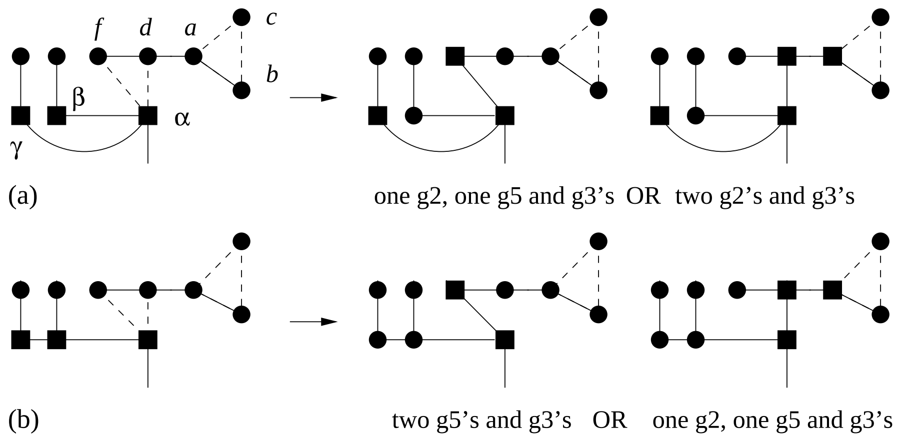


Figure XXX – *cases with two g2's in part (a) of Theorem G.12, when g is a g3*

which in turn is linked to at least one g3 ; then a g4 appears, containing g and δ, and with binding vertices d and δ. So the case when there is a g1 in T' is closed, also completing the case $n = 10$. From now on, we assume that $n \geq 13$ and that there is no g1 in T'.

We can remark the following : if in T' we have a g4, and if one of its binding vertices is connected to a g3, it is always possible to locate $w(T')$ watchers on T' with one watcher at the second binding vertex v of the g4, see Figure XXIX. Since we have required that each of the two b. v.'s of the g4 be connected to at least one b. v. of a g3, this means that we can locate one watcher at one b. v. of the g4 in order to possibly cover α. This remark or Lemma G.8 or G.6 allows us to save one watcher on g when there is one g4 or (at least) one g5 in T', as we did in the left part of Figure XXV(a), with β covering α.

Therefore, we have only one case left when g is a g3 : when T' contains at least one gadget of order 2. If there is exactly one g2 (case (ii) of Theorem G.12), the situation is very close to that of Theorem G.11 (see Figure XVIII and the left part of Figure XVII), the difference being the existence of a g5. So we assume that T' contains two g2's, and g3's (case (i) of Theorem G.12). In general, one can still save one watcher on g ; the two critical situations are given by Figure XXX, in which a single watcher located on g cannot simultaneously cover d, f and β (and γ when γ is connected to α) — cf. Figure XXIV.

• (3) Assume now that g is of order 5, with binding vertex α ($\alpha \neq d$). The other gadgets in T' are either one g2 and $k-3$ g3's, or one g5 and $k-4$ g3's. We illustrate the cases occurring when in T', α is linked to the g2 and only to this gadget : Figure XXXI is for $n = 10$ and uses the only representation with one g5 for a tree of order 7, cf. Figure XXIII(e) ; Figure XXXII is for $n \geq 13$ and is obtained by locating d at all the different vertices, except at the b. v. α, in all the different types for a g5, cf. Figure XX.

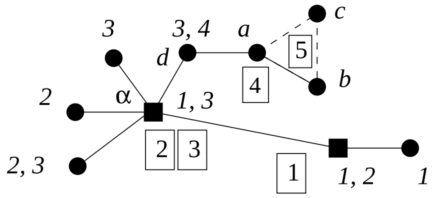


Figure XXXI – *case* $n = 10$ *in part (a) of Theorem G.12, when g is a g5 : five watchers are sufficient*

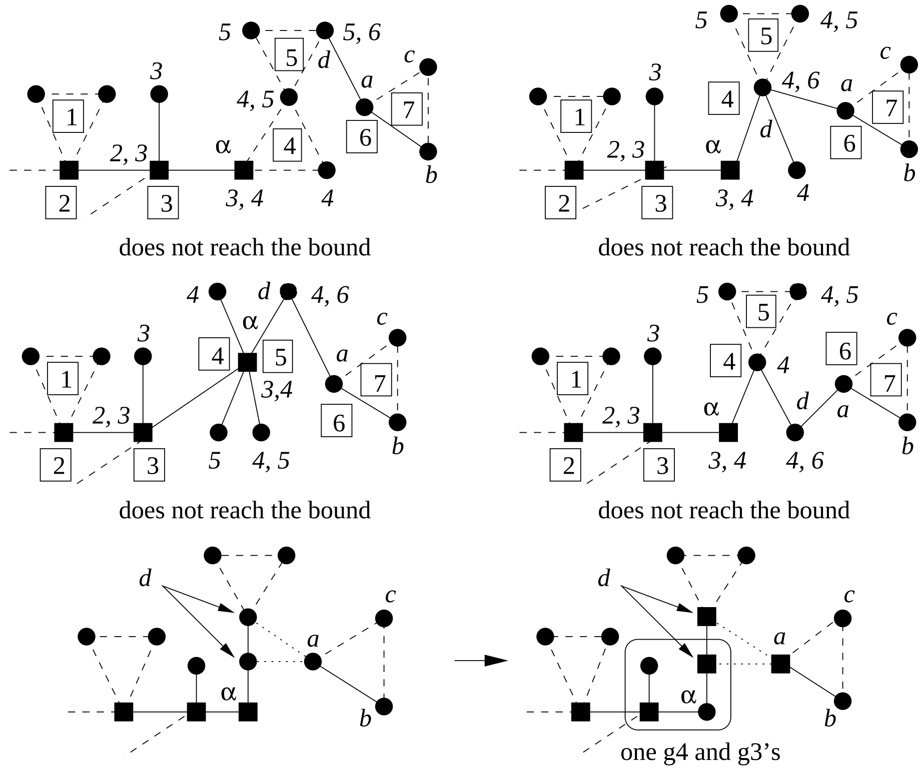


Figure XXXII – *cases when g is a g5 in part (a) of Theorem G.12, for* $n \geq 13$

The other cases, very similar to Figure XXXII or to previous studies involving g5's, often use Lemma G.9 and are left to the reader.

• (4) The final case of this part (a) is when g is a g4, with its two binding vertices α and x ($\alpha \neq d, x \neq d$) connected to other gadgets, and in T' all gadgets except g are g3's (case (v) of Theorem G.12). Denote by T'' the connected component containing x in the forest obtained from T' by removing the edges of g, see Figure XXXIII. Assume that T'' is of order at least 7 ; then by Lemma G.10, x is free in T''. In a minimum watching system for T, we can assume that there is one watcher located at a which covers d, and one watcher located at β which covers α. Then (see Figure XXXIV), either only one more watcher, denoted by 3, is necessary inside g to watch T, and T does not reach the bound $2k$, or T consists of one g2, one g5 and g3's, or of two g5's and g3's. The same argument with α shows that we can assume that each b. v. of g is linked to exactly one g3 ; then $n = 13$ and Figure XXXV gives all the possible cases.

• *Part (b) : we assume that we are in the third or fourth case in the proof of Theorem G.4*

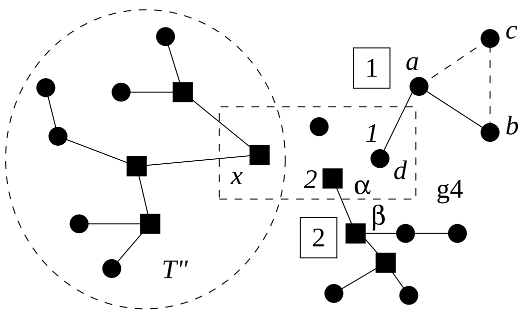


Figure XXXIII – *the definition of T'' in part (a) of Theorem G.12*

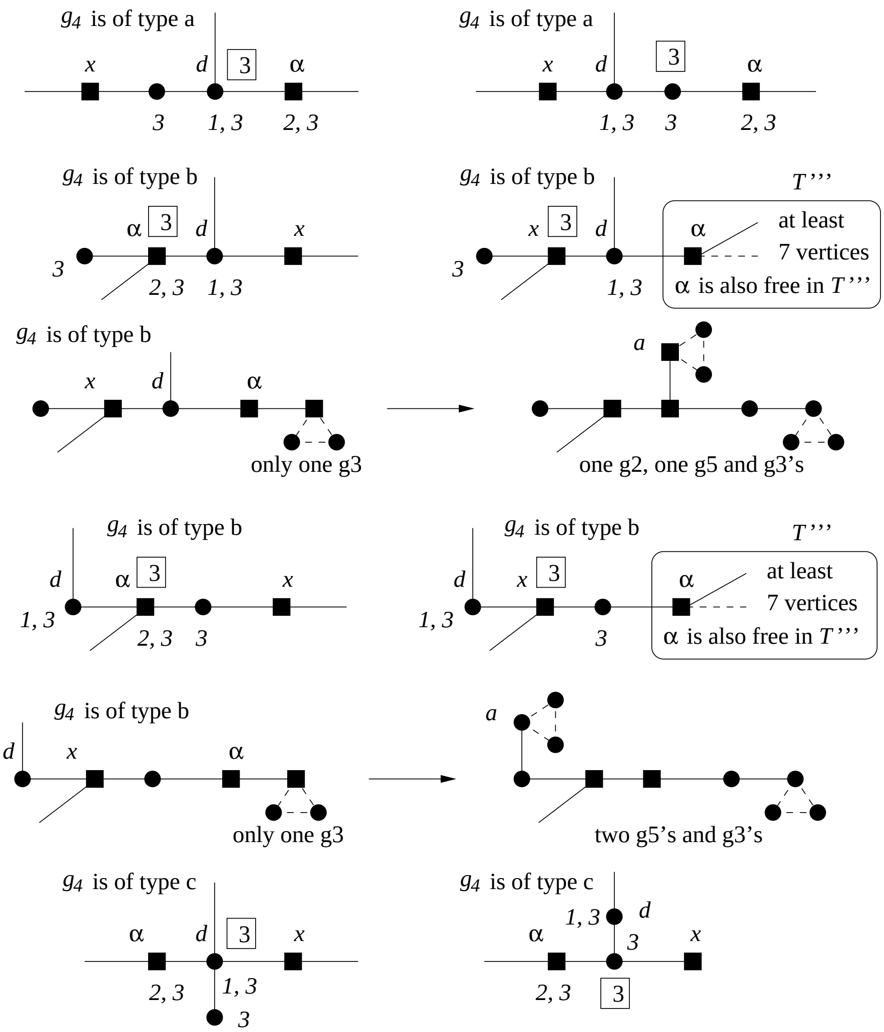


Figure XXXIV – *part (a) of Theorem G.12 : T'' is of order at least 7 and x is free in T''*

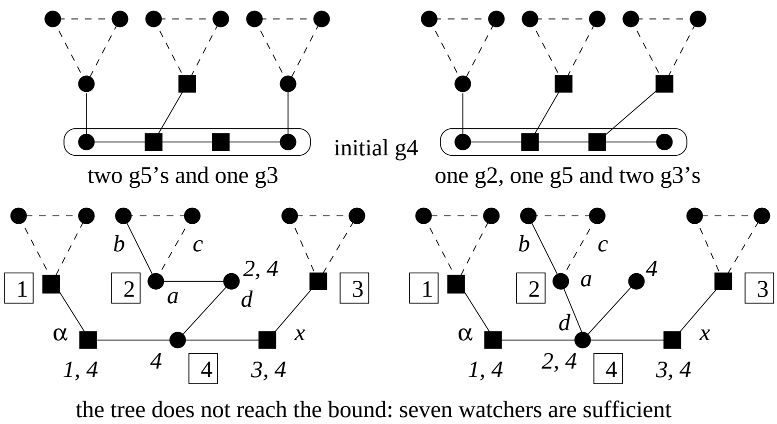


Figure XXXV – *the remaining cases when g is a g4 in part (a) of Theorem G.12*

If we are in the third case, we rename by a, b and d respectively, the vertices x, y and v_{D-1}, and if we are in the fourth case, we rename by a, b and d respectively, the vertices v_{D-1}, v_D and v_{D-2} ; in both cases, we remove a and b, obtaining a tree T' of order $3(k-1)+2$. By Remark G.5, $w(T') = 2(k-1)+1$ and Theorem G.11 may be used : T' can be obtained by choosing one g2 or one g5 and a collection of g3's linked by their binding vertices. Note that the vertex d has degree at least 2 in T' ; it belongs to a gadget g with b. v. α.

We first assume that $d \neq \alpha$. Because of the degree of d, the gadget g cannot be of order 2, and if it is of order 3, with vertex set $\{\alpha, d, c\}$, then its edge set is $\{\{d, \alpha\}, \{d, c\}\}$, and $\{a, b, c, d, \alpha\}$ is a g5 of type c or d : T is of the desired form. We are left with the case when g is a g5, in which d has degree 2 or more, and the other gadgets in T' are all g3's ; this is depicted in Figure XXXVI, where we give the locations of the watchers showing that T does not reach the bound $2k$, or show the b. v.'s of the gadgets involved ; note that if $n \geq 13$, then by Lemma G.10, α is free in T' deprived of the four ordinary vertices of g.

Finally, if $d = \alpha$, then Figures XXXVII–XXXIX give the different cases, according to the order of g. This completes the proof of Theorem G.12. □

G.4 Graphs G reaching the maximum value of $w(G)$

We first give the following definition.

Définition G.2. *A connected graph G is said to be* maximal *if, when we add any edge to G, we obtain a graph G' verifying : $w(G') < w(G)$.*

We denote by $\omega(n)$ the maximum minimum number of watchers needed in a connected graph of order n, i.e.,

$$\omega(n) = \max\{w(G) : G \text{ connected of order } n\}.$$

In the previous section, we have established that $\omega(n) = \lfloor \frac{2n}{3} \rfloor$ for $n \notin \{1, 2, 4\}$, and we have characterized the trees of order n reaching $\omega(n)$. In this section, we want to describe all the maximal connected graphs of order n which reach $\omega(n)$. Using Lemma G.1, the graphs of order n which reach $\omega(n)$ are exactly the connected partial graphs of the maximal connected graphs of order n reaching $\omega(n)$.

We recall that K_p denotes the complete graph (or clique) of order p. Again, we divide our study into three cases, $n = 3k$, $n = 3k+2$ and $n = 3k+1$.

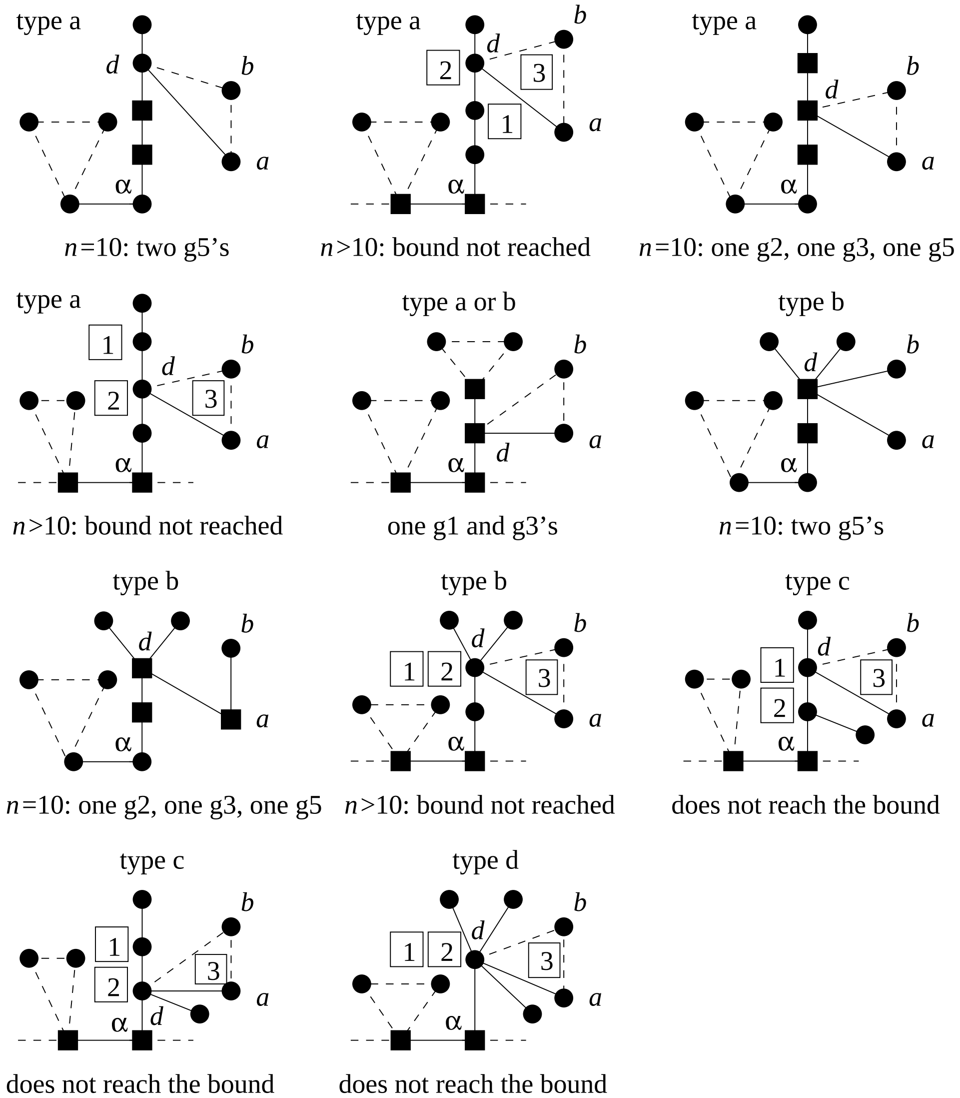


Figure XXXVI – *illustration for part (b) of Theorem G.12, when g is a g5 and $d \neq \alpha$*

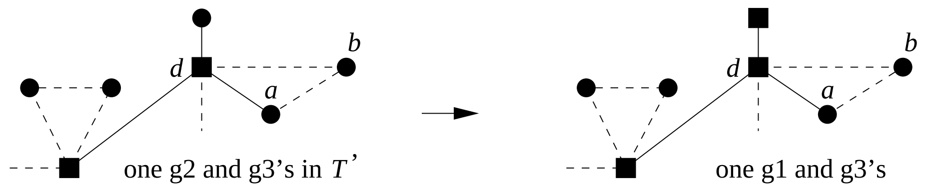


Figure XXXVII – *illustration for part (b) of Theorem G.12, when $d = \alpha$ and g is a g2*

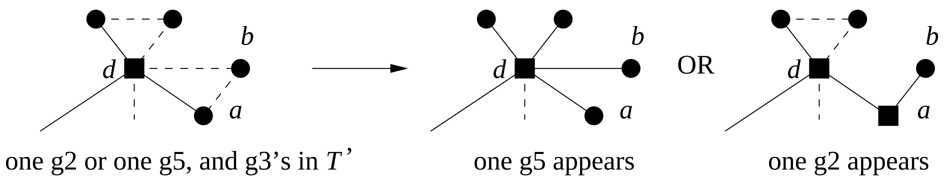


Figure XXXVIII – *illustration for part (b) of Theorem G.12, when $d = \alpha$ and g is a g3*

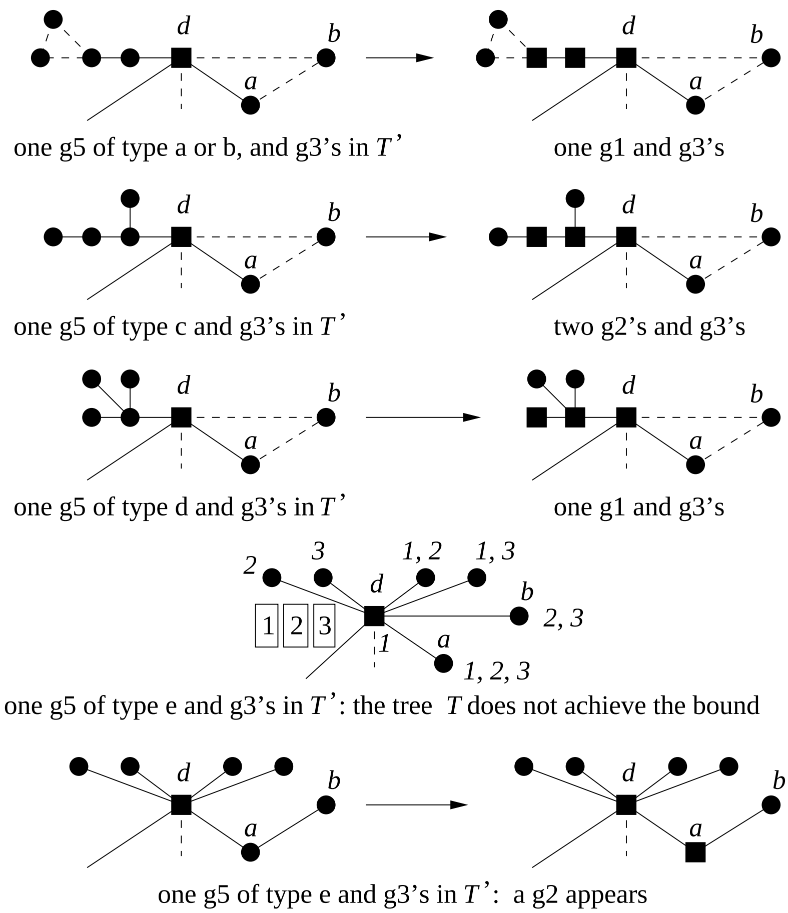


Figure XXXIX – *illustration for part (b) of Theorem G.12, when $d = \alpha$ and g is a g5*

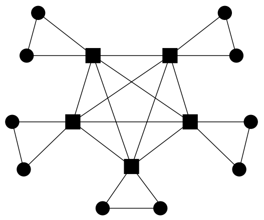


Figure XL – *G_{15}, the maximal graph of order 15 reaching the bound*

Theorem G.13. *Let k be an integer, $k \geq 1$, and G be a maximal graph of order $3k$. We have :*
$w(G) = 2k \Leftrightarrow G$ is obtained by taking a collection of k K_3's, choosing one vertex named a binding vertex *in each K_3, and connecting these k binding vertices by K_k.*

For instance, the graph G_{15} of Figure XL is the unique maximal graph of order 15 reaching the bound $\omega(15) = 10$.

Proof. The implication from the right to the left is direct. So, given a maximal graph G of order $3k$ satisfying $w(G) = 2k$, we have to prove that G is of the form described in the theorem. Let T be a spanning tree of G. Using Lemma G.1 and Theorem G.4, we can see that $w(T) = 2k$. By Theorem G.7, T is a collection of k gadgets of order 3 connected by their binding vertices. We shall show that in G any edge which is not in T is an edge between two b. v.'s of T, or is the missing edge of a g3 ; to do this, we assume that there is in G an edge e which is not an edge between two b. v.'s of T, nor the missing edge of a g3. In Figure XLI, we consider the four possibilities :

(a) The edge e links an ordinary vertex a of a g3, denoted by g_3, whose b. v. is denoted by β, and the b. v. α of another g3, and the edge $\{\alpha, \beta\}$ exists ; then, whatever the type of g_3, we can locate a watcher 3 on g_3 covering a, b and α, and the six vertices are covered and separated by three watchers only.

(b) e links two ordinary vertices of two g3's which are linked by their b. v.'s. Again, the six vertices involved can be watched by three watchers.

In passing, these two cases show how to handle the case $n = 6$, so from now on we assume that $n \geq 9$.

(c) e links an ordinary vertex of a g3, whose b. v. is β, and the b. v. α of another g3, and $\{\alpha, \beta\}$ does not exist. Then α and β are linked to at least one other g3 (possibly the same), because in the spanning tree T, there is a connexion between any two b. v.'s.

(d) This is also true when e links two ordinary vertices of two g3's which are not linked by their b. v.'s.

In each of these two cases, we can see that we are able to locate only one watcher on a g3, so there is a contradiction with the value of $w(G)$.

Furthermore, if we add to T the missing edge on each g3 and all the missing edges between the b. v.'s of T, the number of needed watchers remains equal to $2k$: we have obtained the unique maximal graph containing T. □

Theorem G.14. *(a) Let k be an integer, $k \geq 3$, and G be a maximal graph of order $3k + 2$. We have :*
$w(G) = 2k + 1 \Leftrightarrow G$ is obtained by taking a collection of k K_3's and one K_2, or $k - 1$ K_3's and one K_5, choosing one vertex named a binding vertex in each of these complete graphs, and connecting these binding vertices by K_{k+1} if we have taken a K_2, and by K_k if we have taken a K_5.

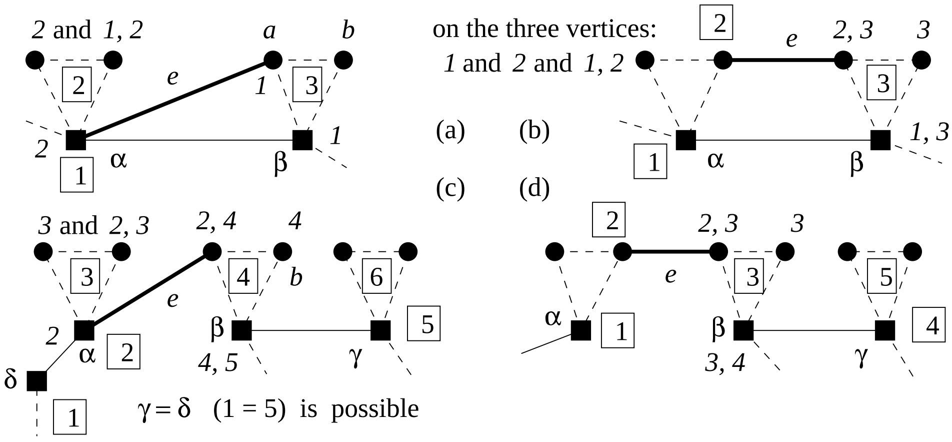


Figure XLI – *forbidden edges between two g3's in the proof of Theorem G.13*

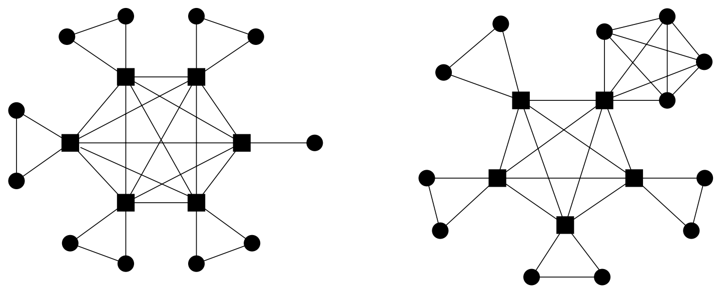


Figure XLII – *the two maximal graphs of order 17 reaching the bound*

(b) If G is a maximal graph of order 8, then we have :
$w(G) = 5 \Leftrightarrow$ *G is the graph given by Figure XLIII, or G is obtained by following the rules given in Case (a), for $k = 2$.*

(c) The only maximal graph G of order 5 with $w(G) = 3$ is the clique K_5.

For instance, the graphs G_{17} and G'_{17} of Figure XLII are the two maximal graphs of order 17 reaching the bound $\omega(17) = 11$.

Proof. The implications from the right to the left are direct. So, given a maximal graph G of order $3k+2$ satisfying $w(G) = 2k+1$, we have to prove that G is of the form(s) described in the theorem.

By inequality (G.1) from the Introduction and Theorem G.4, all connected graphs G of order 5 are such that $w(G) = 3$, K_5 is the unique maximal graph of order 5, and Case (c) is true.

The case $n = 8$, which does not fit the general framework either, is rather tedious to check, and is not given here.

We assume from now on that $n \geq 11$. Let T be a spanning tree of G. Using Lemma G.1 and Theorem G.4, we can see that $w(T) = 2k+1$. From Theorem G.11, T can be obtained

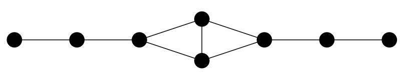


Figure XLIII – *a maximal graph of order 8 reaching the bound*

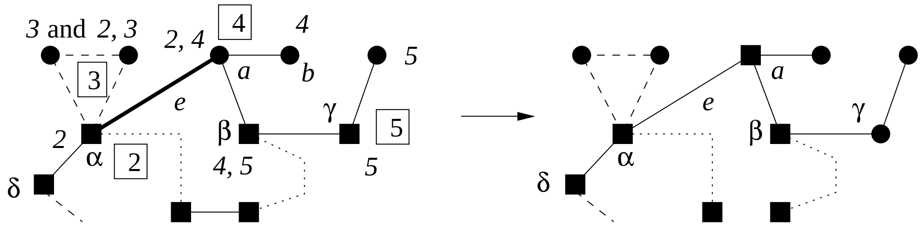


Figure XLIV – *forbidden edges between two g3's : a critical case in part (1) of Theorem G.14*

as one g2 or one g5 plus a collection of g3's, with the gadgets connected by their binding vertices to form a tree. If, among the spanning trees of G, there is one with a g5, we choose this tree; and if, in all the spanning trees, we cannot avoid a g2, then we choose a tree in which the b. v. of the g2 has maximum degree (in the tree).

We shall list pairs of vertices which cannot be adjacent in the maximal graph G : between g3's, between the g5 and a g3, and between the g2 and a g3 (the most delicate case).

• (1) Assume first that there is an edge between two g3's, with at least one of its ends different from a b. v. This case has been treated for Theorem G.13, cf. Figure XLI. If now β, the b. v. of g_3, is not linked to any b. v. other than α, or if β is linked to the b. v. of a g3 other than α, then we can save one watcher in exactly the same way as on Figure XLI. If β is linked to the b. v. γ of the g5, by Lemma G.8 we can have a watcher located at γ and covering β, thus still saving one watcher on g_3. So we can assume that β is linked to the b. v. γ of the g2. In cases (b) and (d) of Figure XLI, we can save one watcher on the g3 with b. v. α, since α and β play symmetrical parts. In case (c), in all cases, but one, we can still save one watcher on g_3 : the critical case (see Figure XLIV) is when the g2 has no connexion other than β, and moreover the watcher 4, which is used to cover b, cannot be located at β, so that the two vertices of the g2 are not separated, and we cannot save one watcher; in this case however, since the b. v.'s $\alpha, \beta, \gamma, \delta, \ldots$ in T are connected, it is possible to add in T the edge $e = \{\alpha, a\}$ and delete one edge between two b. v.'s, so that the result is a spanning tree of G, in which γ becomes an ordinary vertex in a g3, and a becomes the b. v. of the g2, now connected to two g3's. This means that the spanning tree in the left part of the figure cannot have been chosen, since the b. v. of its g2, γ, does not have maximum degree among the spanning trees of G. Case (a) of Figure XLI can be dealt with in the same way, with a critical situation similar to Figure XLIV, where we can add the edge $\{\alpha, a\}$ and delete the edge $\{\alpha, \beta\}$.

• (2) Assume next that there is one g5, named g_5, in T, and that there is in G an edge e between g_5 and a g3, g_3, with at least one of its ends different from a b. v. Let α and β be the b. v.'s of g_5 and g_3, respectively. If the edge $\{\alpha, \beta\}$ does not exist, then α and β are connected to at least one other g3 (possibly the same), and we can save one watcher, using in particular Lemma G.8 : see Figure XLV, where a can be equal to α in the left part. In the right part, since α is free in the tree T' consisting of the spanning tree T deprived of the four ordinary vertices of g_5 (even if α is linked only to γ, in which case α is covered by the watcher 5), we are left with the problem of taking care of the three vertices x, y, z of g_5 other than a and α, with only two watchers; this can be done using Lemma G.9.

So from now on we assume that we have the edge $\{\alpha, \beta\}$ in T. Because $n \geq 11$, α is still free in T', and obviously, if both α and β are still connected to other g3's, the argument above still works. So we assume that only one of α and β is connected to (at least) one

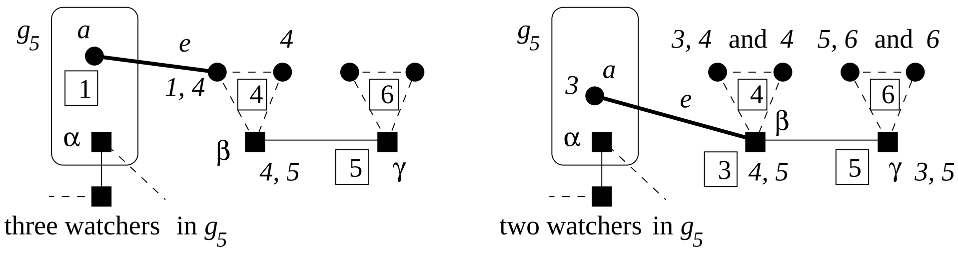


Figure XLV – *part (2) of Theorem G.14 : forbidden edges between a g5 and a g3*

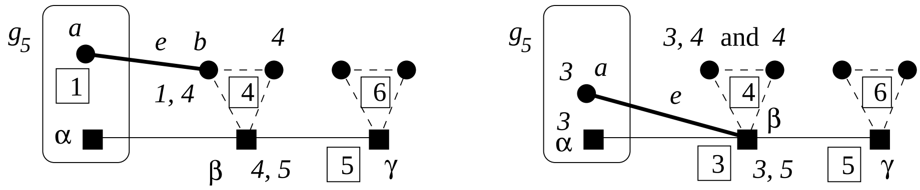


Figure XLVI – *part (2) of Theorem G.14 : forbidden edges between a g5 and a g3, with $\{\alpha, \beta\}$ in T*

(other) g3. We first consider the case when it is β.

Figure XLVI depicts the situation, where a can be equal to α in the left part. In this left part, thanks to Lemma G.8, the situation is the same as previously (without the edge $\{\alpha, \beta\}$). And if e links a and β (see the right part), then, still denoting by x, y and z the vertices in g_5 other than a and α, we can use Lemma G.9 : two watchers are sufficient to watch $\{x, y, z\}$ and cover a, so that a and α are now separated by a watcher. Thus, one watcher can be saved on g_5. We can now assume that β is linked to no gadget other than g_5.

Then the situation is described by Figure XLVII, with α free in T', and $b \neq \beta$ or $b = \beta$ in the left part (in the latter case, locate the watcher 3 at β). When a is one extremity of e, we use Lemma G.9 and save one watcher on $g5$, so we are left with the case $e = \{\alpha, b\}$ with $b \neq \beta$ (see the right part of the figure), which is solved also using Lemma G.9 and saving one watcher on g_3.

• (3) We finally study the case when there is one g2, named g_2, with b. v. α and ordinary vertex α', in the spanning tree T. The situation is now slightly different from the previous cases, because we may, without contradiction, have in G an edge between, for instance, α' and a vertex of a g3, since T may have been originally produced from K_5 in G.

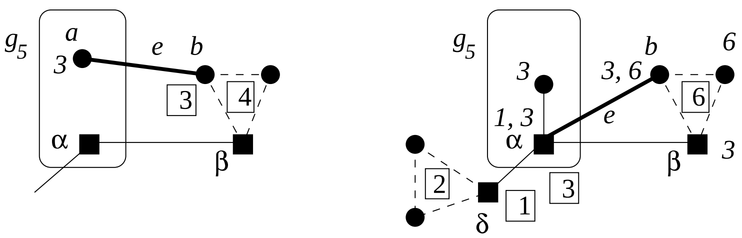


Figure XLVII – *part (2) of Theorem G.14 : more forbidden edges between a g5 and a g3, with $\{\alpha, \beta\}$ in T*

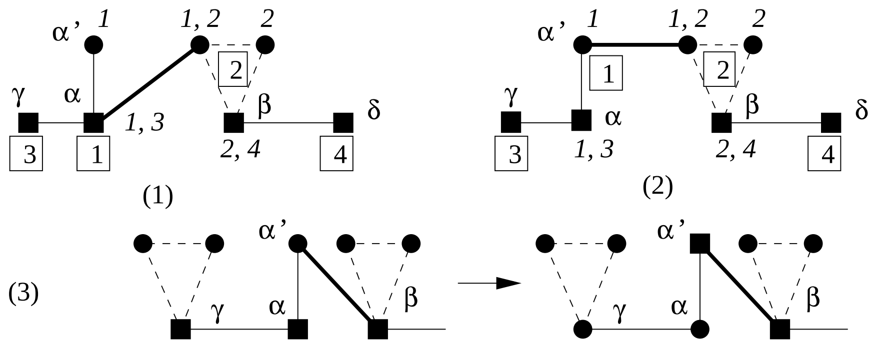


Figure XLVIII – *part (3) of Theorem G.14 : forbidden edges between a g2 and a g3*

We consider in T a $g3$ named g_3, with b. v. β, and investigate which edge(s) can exist in G between g_2 and g_3.

First, we assume that $\{\alpha, \beta\}$ is not in T. Then in T, α is linked to the b. v. γ of a g3, and β is linked to the b. v. δ of a g3, possibly with $\gamma = \delta$. Using Lemma G.6, we locate watchers at γ and δ, and Figure XLVIII(1)(2) shows how to routinely save one watcher on g_3 when in G there is an edge between α or α' and an ordinary vertex of g_3, even if the watchers 3 and 4 coincide. Assume now that it is the edge $\{\alpha', \beta\}$ which is in G. If in T, neither α nor γ is connected to any g3, we are in case (3) of Figure XLVIII and we consider that there is a g5 of type a or b in T rather than a g2. So either α or γ is connected to a g3, with b. v. ϕ. If $\gamma \neq \delta$, then $\phi = \delta$ is possible, or (if ϕ is not linked to α) $\phi = \beta$; if $\gamma = \delta$, then $\phi = \beta$ is possible (if ϕ is not linked to α). All this is depicted in Figure XLIX, where it can easily be seen how to save one watcher on g_2 in all cases ; we give only the full description of the last case, (e).

So we have just established that if α is not connected to the b. v. of a g3, then there exists no edge between this g3 and g_2 in G. What happens now if α is connected to the b. v. β of a g3, g_3, that is to say if there is the edge $\{\alpha, \beta\}$ in T ? If in T, α is still linked to the b. v. γ of a g3 (with $\gamma \neq \beta$) and β is still linked to the b. v. δ of a g3 (with $\gamma \neq \delta$ because there is no cycle in T), we can re-run the argument used in the absence of $\{\alpha, \beta\}$: the first two cases of Figure XLVIII are exactly the same with or without $\{\alpha, \beta\}$, and in the third case, we have the edges $\{\alpha', \beta\}$ and $\{\alpha, \beta\}$ in G, from which we can still pick a spanning tree with a g5 ; and, because $\phi \neq \beta$ and $\phi \neq \delta$, Figure XLIX reduces to its first case (a), which can be treated similarly. Therefore, in T, either α is not linked to the b. v. of any g3 other than β, or β is not linked to the b. v. of any g3.

If in T, α is linked to β only, then we have seen that no edge exists in G between g_2 and any g3 other than g_3. But edges can exist between g_2 and g_3, and indeed, we can add all the missing edges between these two gadgets, plus the missing edge in each g3, plus all the missing edges between the b. v.'s of T, the number of needed watchers remains equal to $2k + 1$, and we have obtained the only maximal graph containing T, which is of the form described in the theorem ; see Figure L. Note that in some cases, the argument of the choice of a g5 in T can also be used, for instance if g_3 is of type b and there is the edge $\{a, \alpha'\}$.

If β is linked only to α and if β is the only b. v. which is linked only to α, then any g3 other than g_3, with b. v. γ, can be linked to g_2 uniquely through the edge $\{\alpha, \gamma\}$, and so in G, the possible edges between the ordinary vertex α' of g_2 and a g3 must affect g_3 only. In the first two cases in Figure LI, a g5 should have been taken when choosing T, or, as in the third case and as in the previous figure, we can add all the edges between g_2

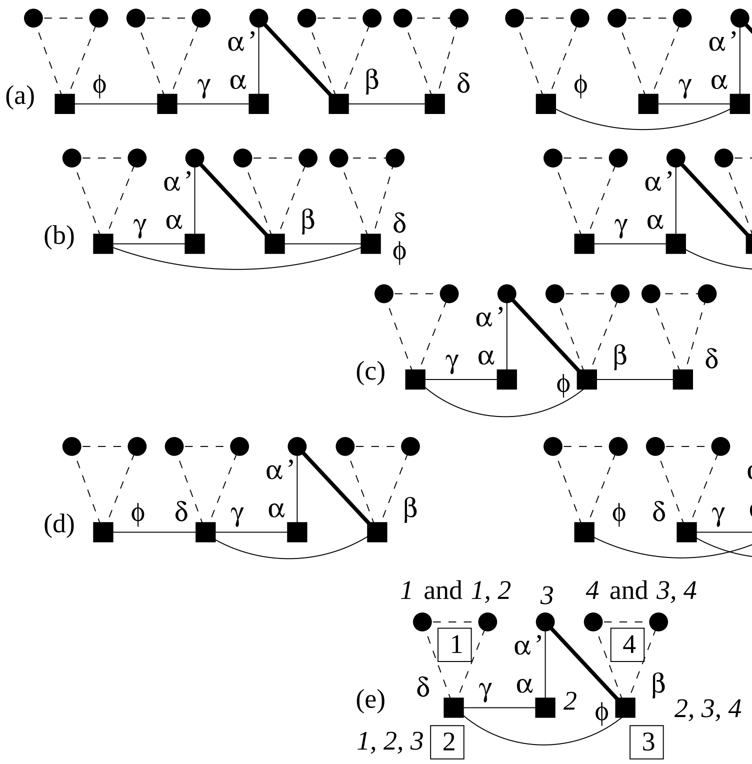


Figure XLIX – *part (3) of Theorem G.14, with the edge* $\{\alpha, \beta\}$ *not in* T *: (a)* $\gamma \neq \delta$*,* $\phi \neq \delta$*,* $\phi \neq \beta$ *; (b)* $\gamma \neq \delta$*,* $\phi = \delta$ *; (c)* $\gamma \neq \delta$*,* $\phi = \beta$ *; (d)* $\gamma = \delta$*,* $\phi \neq \beta$ *; (e)* $\gamma = \delta$*,* $\phi = \beta$*.*

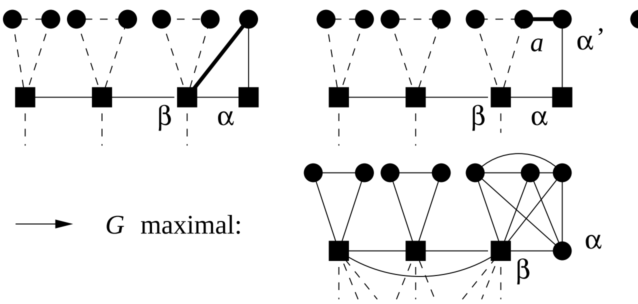


Figure L – *part (3) of Theorem G.14, possible edges between a g2 and a g3 : actually, edges inside* K_5

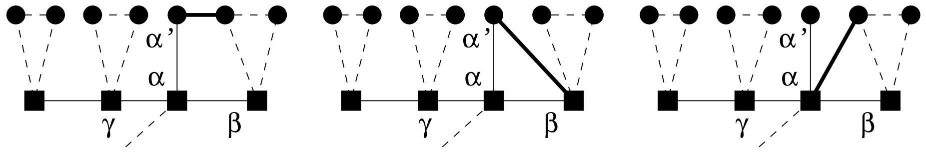


Figure LI – *part (3) of Theorem G.14 : β is the only binding vertex linked only to α*

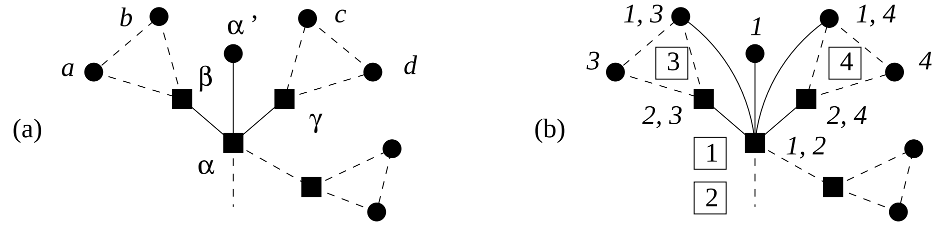


Figure LII – *part (3) of Theorem G.14 : the binding vertices of (at least) two g3's are connected only to α*

and g_3 and obtain K_5. So we are left with the case when there are two (or more) g3's with b. v.'s linked only to α in T, see Figure LII(a). If in G there is the edge $\{\alpha', a\}$, $\{\alpha', b\}$ or $\{\alpha', \beta\}$, then again a spanning tree with a g5 could have been chosen, and if there is the edge $\{\alpha, a\}$ or $\{\alpha, b\}$ and neither $\{\alpha, c\}$ nor $\{\alpha, d\}$, we can add to T all the edges between g_2 and g_3 in order to obtain K_5 in a maximal graph. So the only possibility not ruled out yet is if there are the edges, say, $\{\alpha, b\}$ and $\{\alpha, c\}$ (more edges in G can only help). Then Figure LII(b) shows how to save (at least) one watcher, by locating two watchers at α.

Now we are in a position to conclude. If in T there are edges between ordinary vertices of different gadgets or between the b. v. of a gadget and an ordinary vertex of another gadget, then another spanning tree should have been chosen, containing a g5 instead of a g2, or containing a g2 with binding vertex of higher degree, or these edges are part of K_5, or we can save watchers.

Furthermore, if we add to T the missing edge on each g3, the missing edges on the possible g5, and all the missing edges between the b. v.'s in T, the number of needed watchers remains equal to $2k+1$: we have obtained the only maximal graph containing T. The proof of Theorem G.14 is completed. □

The proof of the previous theorem, for $n = 3k+2$, is not very encouraging in view of the case $n = 3k+1$. Indeed, although we have some insight into the situation, we can only conjecture the following result, in which, to describe the graphs, we need three new gadgets of order 7 (which are not trees), with one or two binding vertex(ices), see Figure LIII.

Conjecture G.3. *Let k be an integer, $k \geq 6$, and G be a maximal graph of order $3k+1$. We have :*

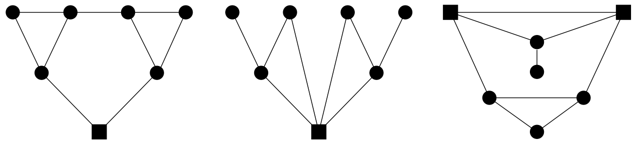


Figure LIII – *the three new gadgets of order 7*

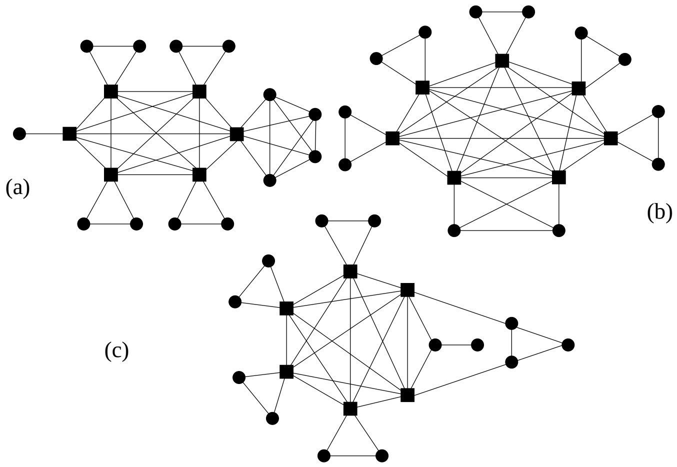


Figure LIV – *three graphs reaching the bound* $\omega(19) = 12$

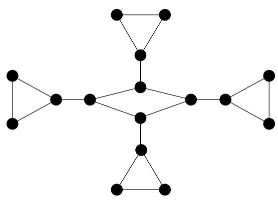


Figure LV – *a maximal graph reaching the bound* $\omega(16) = 10$

$w(G) = 2k \Leftrightarrow G$ is obtained by :

- (i) *taking two K_2's and $k-1$ K_3's,*
- (ii) *or taking one K_2, one K_5 and $k-2$ K_3's,*
- (iii) *or taking two K_5's and $k-3$ K_3's,*
- (iv) *or taking one K_4 and $k-1$ K_3's,*
- (v) *or taking one g7 and $k-2$ K_3's,*

choosing one vertex named a binding vertex on each of the complete components K_i, except on K_4 for which we choose two binding vertices, taking for the g7 one or two binding vertex(ices) according to its type, and connecting these binding vertices to form a complete graph with them.

The graphs of Figure LIV are graphs of order 19 reaching the bound $\omega(19) = 12$: (a) with one K_2, one K_5 and four K_3's ; (b) with one K_4 and five K_3's ; (c) with one g7 and four K_3's ; according to Conjecture G.3, they would be maximal.

For $n = 3k + 1$ with $k \leq 5$, there are maximal graphs needing $2k$ watchers which are not of the form described in the conjecture. We give a certified example for $n = 16$ in Figure LV.

Annexe H

On the Sizes of G, G^r, $G^r \setminus G$, Part I : the Undirected Case

David Auger [1], **Irène Charon** [1],
Olivier Hudry [1], **Antoine Lobstein** [2]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Abstract

Let G be an undirected graph and G^r be its r-th power. We study different issues dealing with the number of edges in G and G^r. In particular, we answer the following question : given an integer $r \geq 2$ and a connected graph G of order n such that $G^r \neq K_n$, what is the minimum number of edges that are added when going from G to G^r, and which are the graphs achieving this bound ?

Keywords : Graph Theory, Undirected Graph, Diameter, Power of a Graph.
2000 *Mathematics Subject Classification :* 68Q17, 05C99, 94B65.

H.1 Introduction

We first give some very basic definitions and notation for undirected graphs, before we expound our study.

H.1.1 Definitions and Notation

We shall denote by $G = (V, E)$ an *undirected graph* with vertex set V and edge set E, where an *edge* between $x \in V$ and $y \in V$ is indifferently denoted by the sets $\{x, y\}$ or $\{y, x\}$, x and y being called the *extremities* of the edge. We require the graph to have no loops nor double edges. The *size* of an undirected graph is its number of edges, its *order* is its number of vertices.

2. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

An undirected *path* $P = x_0x_1 \dots x_\ell$ is a sequence of vertices x_i, $0 \le i \le \ell$, such that $\{x_i, x_{i+1}\} \in E$ for i between 0 and $\ell - 1$. The *length* of P is its number of edges, ℓ. An undirected graph is called *connected* if for any two vertices x and y there is a path between x and y.

In a connected graph G, we can define the distance between any vertex x and any vertex y, denoted by $d_G(x, y)$, as the number of edges in any shortest undirected path between x and y, since such a path always exists. Note that we always have $d_G(x, y) = d_G(y, x)$. The *diameter* of a connected graph G is the maximum distance in the graph :

$$\mathrm{diam}(G) = \max_{x \in V, y \in V} d_G(x, y).$$

Given an integer $r \ge 1$, the *r-th power*, or *r-th transitive closure*, of the graph $G = (V, E)$ is the graph $G^r = (V, E^r)$, where, for two distinct vertices x and y, the edge $\{x, y\}$ is in E^r if and only if $d_G(x, y) \le r$.

The *clique*, or *complete graph*, K_n, is the graph of order n with all possible $n(n-1)/2$ edges. Finally, the *subgraph of $G = (V, E)$ induced by $V^* \subseteq V$* is the graph $G^* = (V^*, E^*)$ where $E^* = \{\{x, y\} : x \in V^*, y \in V^*, \{x, y\} \in E\}$.

H.1.2 Scope of the Paper

We are interested in the following related problems on sizes and powers, for undirected graphs :

(I) Given the order and diameter of a connected graph, what is its maximum size, and which are the graphs achieving this bound ? This issue was completely solved by Ore back in 1968, see Section H.1.3 below.

(II) Given an integer $r \ge 2$, what is the minimum size of a graph of order n, of which it is known that it is the r-th power of a connected graph, and which are the graphs achieving this bound ? The minimum size is already known and given in Theorem H.2, and we easily characterize the graphs which achieve it in Section H.2.1.

(III) Given an integer $r \ge 2$,

(i) given a connected graph G of order n such that $G^r \ne K_n$, what is the minimum number of edges that are added when going from G to G^r ? The answer is known for $r = 2$, see Theorem H.3 ;

(ii) which are the graphs achieving this bound ?

(iii) how many edges can we have in the graphs reaching the bound ? This is Question 1, formulated for $r = 2$, in [1].

We give the complete answer to question (III) in Section H.2.2, which is the core of our article.

The same issues for directed graphs are treated in [10].

H.1.3 Known Results

(I) As aforementioned, the maximum size of an undirected graph of given order and diameter is known, as well as the families of graphs which achieve this bound.

Theorem H.1. *[79, Th. 3.1] Let $G = (V, E)$ be an undirected connected graph of order n and diameter $\delta \ge 2$. Then the size of G is at most*

$$\delta + \frac{(n - \delta - 1)(n - \delta + 4)}{2}.$$

Proof. We give a very simple proof, different from that by Ore. Let $z_1, z_2 \in V$ be such that $d_G(z_1, z_2) = \delta$, and C be a shortest path between them : $C = x_0x_1 \dots x_\delta$, with $x_0 = z_1$ and $x_\delta = z_2$, and there are no more edges between vertices x_i. In G, the remaining vertices y_j, $1 \leq j \leq n - \delta - 1$, can at most constitute the clique $K_{n-\delta-1}$, and each y_j can be part of at most three edges with extremities in C : this is obvious if $\delta = 2$, and if $\delta \geq 3$, there would otherwise be two edges $\{y_j, x_{i_1}\}$ and $\{y_j, x_{i_2}\}$ with $i_1 + 3 \leq i_2$, and the path $x_0 \dots x_{i_1} y_j x_{i_2} \dots x_\delta$ would be shorter than C. Summing up, we have at most

$$\delta + \frac{1}{2}(n - \delta - 1)(n - \delta - 2) + 3(n - \delta - 1)$$

edges in G. □

We shall see that this theorem is also a direct consequence of Theorem H.6. We set

$$s(\delta, n) = \delta + \frac{(n - \delta - 1)(n - \delta + 4)}{2};$$

the graphs achieving $s(\delta, n)$ are characterized in [79, Sec. 3]. The reader will easily understand how to do this after reading the slightly more complicated case of directed graphs, in [10, Sec. 2].

(II) The following theorem is about the size of the power of a graph.

Theorem H.2. *[22, Sec. 9.3] If $G = (V, E)$ is a connected graph of order n and if $r < diam(G)$, then the size of G^r is at least*

$$nr - \frac{r(r+1)}{2}, \tag{H.1}$$

and this bound is achieved by the path $x_0x_1 \dots x_{n-1}$. □

We shall see in Section H.2.1 that the path is the only graph achieving the lower bound (H.1).

(III) Next theorem gives the minimum number of edges added when going from G to G^2.

Theorem H.3. *[1, Th. 1] If G^2 is not a complete graph, then the number of edges in $E^2 \setminus E$ is at least $n - 2$.* □

The authors also exhibit graphs which achieve the bound $n - 2$ [1, Fig. 1], but they do not provide a characterization.

H.2 Our Results

The maximum size of an undirected graph of given order and diameter and the families of graphs achieving the bound having been evoked in Section H.1.3, we go directly to the following problem : which are the graphs achieving the bound given by (H.1) ?

H.2.1 Size of the Power of a Graph

We have seen in Theorem H.2 that the size of G^r is at least $nr - \frac{r(r+1)}{2}$, and we now characterize the graphs which meet this bound.

Theorem H.4. *Let $G = (V, E)$ be a connected graph of order n such that G^r has size exactly $nr - \frac{r(r+1)}{2}$, with $r < diam(G)$.*
Then G is the path $P_n = x_0x_1 \dots x_{n-1}$.

Proof. It can be done by induction on n. The small cases are easy to handle. The induction assumption is that the set of connected graphs G^* of order $n-1$ which meet the bound $(n-1)r - \frac{r(r+1)}{2}$ for $r <$ diam(G^*) is reduced to $\{P_{n-1}\}$.

We consider a connected graph $G = (V, E)$ of order n such that G^r has size exactly $nr - \frac{r(r+1)}{2}$ for $r <$ diam(G), and a vertex $z_0 \in V$ such that there is a vertex $z_1 \in V$ with $d_G(z_0, z_1) =$ diam(G) ; we let $G^* = (V^*, E^*)$ be the subgraph of G induced by $V^* = V \setminus \{z_0\}$. Because of the choice of z_0, G^* is connected. It is also obvious that in G^r there are at least r edges with z_0 as an extremity. Therefore, because the bound (H.1) is linear in n with the factor r, there are in G^r exactly r edges with z_0 as an extremity, and G^* meets the lower bound $(n-1)r - \frac{r(r+1)}{2}$.

If $r <$ diam(G^*), then we can apply the induction assumption, and G^* is the path $P_{n-1} = x_0x_1 \ldots x_{n-2}$. Then it is straightforward to see that z_0, which must bring exactly r edges to G^r, is linked either to x_0 only or to x_{n-2} only, so G is the path with n vertices.

If $r \geq$ diam(G^*), then $(G^*)^r = K_{n-1}$, and $|(E^*)^r| = (n-1)(n-2)/2$, which is also equal to $(n-1)r - \frac{r(r+1)}{2}$. This leads to $r = n-1$ or $r = n-2$. But $r <$ diam$(G) \leq n-1$, so only $r = n-2$ is possible, and this implies that diam$(G) = n-1$, i.e., G is the path P_n. □

H.2.2 From G to G^r

Given an integer $r \geq 2$ and all connected undirected graphs $G = (V, E)$, of order n, such that $G^r \neq K_n$, what is the minimum number of edges that are added when going from G to G^r, i.e., what is the smallest $|E^r \setminus E|$? we shall denote this number by $\mathcal{A}(r, n)$. Once we have determined $\mathcal{A}(r, n)$, we shall characterize the graphs achieving this number, and study their sizes.

Observe that diam$(G) \geq r + 1$, because $G^r \neq K_n$.

We omit the proof of the following lemma, which uses only basic counting arguments ; we set

$$b(r, n) = (r-1)(n-1-\frac{r}{2}).$$

Lemma H.5. *Let G be a path with n vertices, such that $G^r \neq K_n$. Then exactly $b(r, n)$ edges are added when going from G to G^r.* □

The following theorem answers question (III)(i) of Section H.1.2 and shows in particular that $\mathcal{A}(r, n)$ is linear in n, with the factor $r-1$. When $r = 2$, it returns the value $n-2$ of Theorem H.3.

Theorem H.6.

$$\mathcal{A}(r, n) = b(r, n). \tag{H.2}$$

Proof. Note that equality (H.2) also contains the case $r = 1$. We begin by showing that $\mathcal{A}(r, n) \geq b(r, n)$. Let $\overline{E}$ be the complementary set of edges in any graph $G = (V, E)$:

$$\overline{E} = \{\{x, y\} : x \in V, y \in V, x \neq y\} \setminus E.$$

We first give a lower bound on $|\overline{E}|$ in Lemma H.7, in which we assume that we can partition the vertex set V into $p \geq 3$ subsets V_i, $1 \leq i \leq p$, in the following way :

– there is a path $x_1x_2 \ldots x_p$ with $x_i \in V_i$, $1 \leq i \leq p$;

– there is no edge between V_i and V_j if $|i-j| \geq 2$.

We set $\alpha = |V_1 \setminus \{x_1\}|$, $\beta = |V_p \setminus \{x_p\}|$.

Lemma H.7.

$$|\overline{E}| \geq \frac{(p-1)(p-2)}{2} + (n-p)(p-3) + \alpha + \beta + \alpha\beta. \tag{H.3}$$

Proof of Lemma H.7. In $\overline{E}$, we have the following edges :

(i) the edges $\{x_i, x_j\}$, $1 \le i < j \le p$, $j - i \ge 2$. Their number is

$$\frac{(p-1)(p-2)}{2}\,;$$

(ii) the edges $\{y, x_j\}$, $y \in V_i \setminus \{x_i\}$, $1 \le i, j \le p$, $|j - i| \ge 2$. For a given y, there are $p-3$ such edges if $i \ne 1$ and $i \ne p$, and $p-2$ if $i = 1$ or $i = p$. Adding up for all such y's, we see that the number of edges of type (ii) is

$$(p-2)(\alpha+\beta) + (p-3)(n-p-\alpha-\beta) = (p-3)(n-p) + \alpha + \beta;$$

(iii) the $\alpha\beta$ edges $\{y, z\}$ where $y \in V_1 \setminus \{x_1\}$, $z \in V_p \setminus \{x_p\}$.

Inequality (H.3) immediately follows, since no edge is counted twice. □

Now we start an induction on the order of the graph, for a given r. Since we deal with graphs of diameter at least $r+1$, we start with graphs $G_0 = (V_0, E_0)$ of order $n = r+2$; G_0 is then a path $x_0x_1 \dots x_{r+1}$, and using Lemma H.5, we know that $|E_0^r \setminus E_0| = b(r, r+2)$.

Next, we assume that $|(E^*)^r \setminus E^*| \ge b(r, n-1)$ for all connected graphs $G^* = (V^*, E^*)$ of order $n-1$, with $n-1 \ge r+2$, which are such that $(G^*)^r \ne K_{n-1}$, and we consider a connected graph $G = (V, E)$ of order n such that $G^r \ne K_n$. We set $\delta = \mathrm{diam}(G)$.

We consider two vertices z_1 and z_2 at distance δ from one another in G, and a shortest path between them : $C = x_0x_1 \dots x_\delta$, with $x_0 = z_1$, $x_\delta = z_2$ and $\delta \ge r+1 \ge 3$. Our description of the graph G is viewed from x_0, and we partition V into $\delta+1$ sets V_i, $0 \le i \le \delta$, according to the distances to x_0 : $V_i = \{y \in V : d_G(y, x_0) = i\}$, so that $V_0 = \{x_0\}$ and each V_i contains x_i. Let G^* be the subgraph of G induced by $V^* = V \setminus \{x_0\}$. Note that G^* is connected. We distinguish between two cases.

Case 1, $(G^*)^r = K_{n-1}$. Then obviously $\mathrm{diam}(G^*) = r$ and $\delta = r+1$. We set $\alpha = |V_1 \setminus \{x_1\}|$, $\beta = |V_\delta \setminus \{x_\delta\}|$. Applying Lemma H.7 to G^*, whose vertex set is suitably partitioned into $r+1 = \delta \ge 3$ subsets V_i, $1 \le i \le r+1$, and whose order is $n-1$, we know that there are at least

$$\frac{r(r-1)}{2} + (n-r-2)(r-2) + \alpha + \beta + \alpha\beta \tag{H.4}$$

edges not in G^*, hence these edges belong to $(E^*)^r \setminus E^*$, and also to $E^r \setminus E$.

In $E^r \setminus E$ are also the edges $\{x_0, y\}$, with $y \in V_2 \cup V_3 \cup \dots \cup V_r$, i.e.,

$$n - (1 + (\alpha+1) + (\beta+1)) = n - (\alpha+\beta+3) \tag{H.5}$$

edges. Summing up (H.4) and (H.5), one gets at least $b(r, n) + \alpha\beta$ edges in $E^r \setminus E$. Because $\alpha\beta \ge 0$, the desired lower bound is obtained in Case 1.

Case 2, $(G^*)^r \ne K_{n-1}$. Then the induction hypothesis works : there are in $(E^*)^r \setminus E^*$ at least $b(r, n-1)$ edges. On the other hand, $E^r \setminus E$ contains the $r-1$ edges $\{x_0, x_i\}$, $2 \le i \le r$, and all in all there are at least

$$b(r, n-1) + (r-1) = b(r, n)$$

edges in $E^r \setminus E$, which closes our second case, ends the induction and shows that $\mathcal{A}(r, n) \ge b(r, n)$.

Finally, we exhibit a graph showing that $\mathcal{A}(r, n) \le b(r, n)$. It simply consists of a path $x_0x_1 \dots x_{n-1}$ with $n \ge r+2$ vertices : Lemma H.5 states that we add $b(r, n)$ edges when going from this graph to its r-th power. This ends the proof of Theorem H.6. □

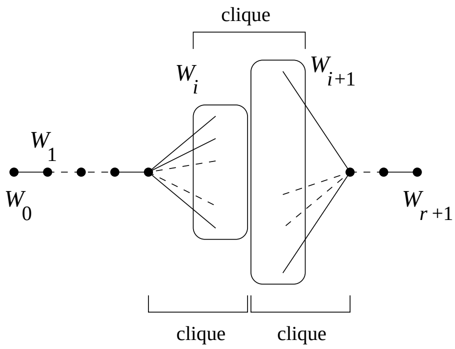


Figure I – *Type 1 graphs in Theorem H.8 : $\delta = r+1$. Any set of vertices represented by a rounded off rectangle may be reduced to a single vertex. The subscript i can vary between* 0 *and r.*

As promised, we show that Theorem H.6 directly gives Theorem H.1 : let $G = (V, E)$ be a connected graph of order n and diameter $\delta \geq 2$; then $G^{\delta-1} \neq K_n$, $|E^{\delta-1}| \leq \frac{n(n-1)}{2} - 1$, and $|E| \leq |E^{\delta-1}| - b(\delta - 1, n)$. Calculations show that $\frac{n(n-1)}{2} - 1 - b(\delta - 1, n) = s(\delta, n)$, including when $\delta = 2$.

We now characterize the graphs of order n which are such that $|E^r \setminus E| = b(r, n)$, which will answer question (III)(ii) in Section H.1.2. We shall say that a graph $G = (V, E)$ is a *sum* $(W_0, W_1, \ldots, W_q)$ *of cliques* if it meets the following conditions :
– the sets $W_0, W_1, \ldots, W_q$ partition V ;
– for i between 1 and q, the subgraph of G induced by $W_{i-1} \cup W_i$ is a clique ;
– for i and j between 0 and q, $|i - j| \geq 2$, there is no edge between W_i and W_j in E.

Note that a graph meeting these conditions has diameter q, and that two vertices $y \in W_i$ and $z \in W_j$, $i \neq j$, are at distance $|i - j|$ from one another.

The graphs described in Theorem H.8 are illustrated in Figures I and II.

Theorem H.8. *Let $G = (V, E)$ be a connected graph of order n such that $G^r \neq K_n$ and $|E^r \setminus E| = b(r, n)$. Then, if diam$(G) = \delta$, G is of one of the following two types :*
– Type 1 : G is a sum $(W_0, W_1, \ldots, W_\delta)$ of cliques such that $\delta = r + 1$ and if, for $0 \leq i < j \leq r + 1$, one has $|W_i| \geq 2$ and $|W_j| \geq 2$, then $j = i + 1$: in other words, there are at most two values of i such that $|W_i| \geq 2$, and if they exist, these two values are consecutive ;
– Type 2 : G is a sum $(W_0, W_1, \ldots, W_\delta)$ of cliques with $|W_i| = 1$ for $2 \leq i \leq \delta - 2$; moreover, if $\delta = r + 2$, then $|W_1| = 1$ or $|W_{\delta-1}| = 1$, and if $\delta = r + 1$, then $|W_1| = |W_{\delta-1}| = 1$.

Proof. First, it has to be checked that these graphs satisfy $|E^r \setminus E| = b(r, n)$. This can be seen using the following argument : first, in all cases there are at least r singleton sets W_i ; second, if all sets W_i are singletons, i.e., we have a path, then we know that the graph meets $b(r, n)$; third, we observe that, starting from a path, if we add one by one the vertices y belonging to the sets W_j which are not singletons, each vertex y brings exactly $r - 1$ new edges in $E^r \setminus E$, namely the edges $\{y, z\}$ with $W_i = \{z\}$ and $r \geq |i - j| \geq 2$; these edges are counted only once ; finally, we use that $b(r, n)$ is linear in n, with the factor $r - 1$.

Now we consider a graph $G = (V, E)$ of order n and diameter δ which meets $b(r, n)$, and show, by induction on n, that G is of type 1 or 2. Since the diameter is at least $r + 1$, we start with $n = r + 2$, in which case the graph is a path, has diameter $r + 1$, and is

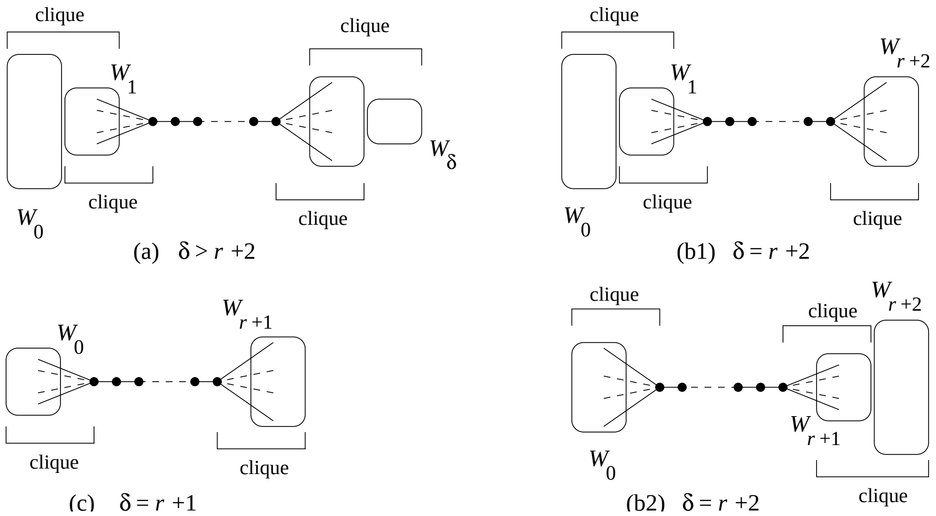


Figure II – *Type 2 graphs in Theorem H.8 : $\delta > r+2$, $\delta = r+2$, or $\delta = r+1$.*

of type 1, or of type 2 with $\delta = r+1$. The induction assumption is that Theorem H.8 is true for $n-1$, $n-1 \geq r+2$. We follow the proofs of Theorem H.6 and Lemma H.7, consider two vertices $z_1 = x_0$ and z_2 at distance δ from one another in G, partition V into $\delta+1$ sets $V_i \supseteq \{x_i\}$ according to the distances to x_0, and denote by G^* the subgraph of G induced by $V^* = V \setminus \{x_0\}$. Again, we have two cases.

Case 1, $(G^*)^r = K_{n-1}$. Then $\text{diam}(G^*) = r$, $\delta = r+1$, and in order to obtain exactly $b(r,n)$ when summing (H.4) and (H.5), necessarily $\alpha\beta = 0$, where $\alpha = |V_1 \setminus \{x_1\}|$, $\beta = |V_\delta \setminus \{x_\delta\}|$. Furthermore, when counting the edges in the proof of Lemma H.7 which give the lower bound (H.4) on the number of edges not in G^*, we did not consider the edges inside the sets V_i, $1 \leq i \leq \delta$, nor between two consecutive sets V_i, V_{i+1}, $1 \leq i \leq \delta-1$; this means that, if we have equality in (H.4), these edges are in G^* : we obtain that G is a sum of cliques $(V_0, V_1, \ldots, V_\delta)$, with $V_0 = \{x_0\}$, and, because $\alpha\beta = 0$, $|V_1| = 1$ or $|V_\delta| = 1$. But in Lemma H.7 we never counted the edges $\{y,z\}$, $y \in V_i \setminus \{x_i\}$, $z \in V_j \setminus \{x_j\}$, $1 \leq i < j \leq \delta$, $j-i \geq 2$, $(i,j) \neq (1,\delta)$. Since these edges belong to $(E^*)^r \setminus E^*$, this shows that such couples y, z do not exist, i.e., $|X_i| \geq 2$ and $|X_j| \geq 2$ is impossible if $|i-j| \geq 2$. So G is of type 1.

Case 2, $(G^*)^r \neq K_{n-1}$. In the proof of Theorem H.6, Case 2 shows that we reach $b(r,n)$ if and only if $|(E^*)^r \setminus E^*| = b(r, n-1)$ and

$$|\{\{x_0, y\} \in E^r \setminus E : y \in V\}| = r-1. \tag{H.6}$$

By the induction hypothesis, G^* is of type 1 or 2. Let δ^* be the diameter of G^* ; $\delta^* = \delta$ or $\delta^* = \delta - 1$, and $\delta^* \geq r+1$. In both types, G^* is a sum $(W_0^*, W_1^*, \ldots, W_{\delta^*}^*)$ of cliques. For i between 0 and δ^*, we choose one vertex x_i^* in each W_i^* ; the path $x_0^* x_1^* \ldots x_{\delta^*}^*$ is in G. The following lemma will be used repeatedly.

Lemma H.9. *(a) The vertex x_0 has no neighbour in W_i^*, $2 \leq i \leq \delta^*-2$.*

(b) The vertex x_0 cannot have a neighbour both in $W_0^* \cup W_1^*$ and in $W_{\delta^*-1}^* \cup W_{\delta^*}^*$.

(c) If x_0 has one neighbour in W_0^* and none in W_1^*, then all the edges between x_0 and W_0^* are in E, and $|W_1^*| = |W_2^*| = \ldots = |W_{r-1}^*| = 1$.

(c') If x_0 has one neighbour in $W_{\delta^*}^*$ and none in $W_{\delta^*-1}^*$, then all the edges between x_0 and $W_{\delta^*}^*$ are in E, and $|W_{\delta^*-1}^*| = |W_{\delta^*-2}^*| = \ldots = |W_{\delta^*-r+1}^*| = 1$.

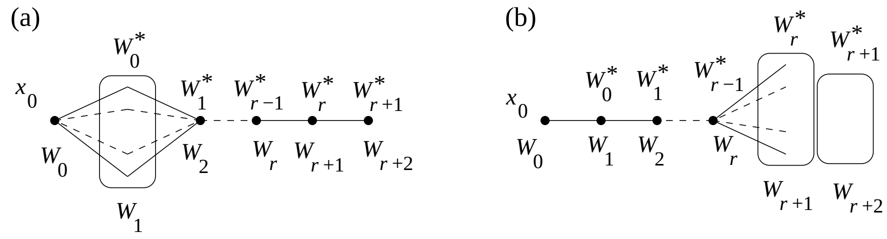


Figure III – *G^* is of type 1, x_0 has no neighbour in W_1^*, and G is of type 2 with $\delta = r+2$; (a) $|W_0^*| \geq 2$; (b) $|W_0^*| = 1$.*

(d) If x_0 has a neighbour in W_1^*, then all the edges between x_0 and $W_0^* \cup W_1^*$ are in E, and $|W_2^*| = \ldots = |W_r^*| = 1$.

(d') If x_0 has a neighbour in $W_{\delta^*-1}^*$, then all the edges between x_0 and $W_{\delta^*}^* \cup W_{\delta^*-1}^*$ are in E, and $|W_{\delta^*-2}^*| = \ldots = |W_{\delta^*-r}^*| = 1$.

Proof of Lemma H.9. (a) Assume that $y \in W_i^*$ exists, with $\{x_0, y\} \in E$ and $2 \leq i \leq \delta^* - 2$. Remember that there is a vertex z_2 at distance δ from x_0 in G. Now $z_2 \in W_j^*$ for some j between 0 and δ^* ; we can assume that $i \neq j$, otherwise $d_G(x_0, z_2) \leq 2$, since W_i^* is a clique. Then, because G^* is a sum of cliques, $d_G(y, z_2) = |i-j|$, whence $d_G(x_0, z_2) \leq |i-j| + 1 \leq (\delta^* - 2) + 1 < \delta$, a contradiction.

(b) Assume that $y_1 \in W_0^* \cup W_1^*$ and $y_2 \in W_{\delta^*-1}^* \cup W_{\delta^*}^*$ exist, with $\{x_0, y_1\} \in E$ and $\{x_0, y_2\} \in E$. Then the vertex $z_2 \in W_j^*$, $0 \leq j \leq \delta^*$, is at distance at most j from y_1 and at most $\delta^* - j$ from y_2, which implies that $d_G(x_0, z_2) \leq \min\{j+1, \delta^* - j + 1\}$, contradicting $d_G(x_0, z_2) = \delta$.

(c) Obviously, the $r-1$ edges $\{x_0, x_i^*\}$, $1 \leq i \leq r-1$, are in E^r, and they are not in E, thanks to the assumption for $i = 1$, and to Lemma H.9(a) and (b) for $2 \leq i \leq r-1$; consequently, by (H.6) there is no other edge $\{x_0, y\}$ in $E^r \setminus E$, which implies that all the edges between x_0 and W_0^* are in E, and that $|W_1^*| = \ldots = |W_{r-1}^*| = 1$. The proof of (c') is similar.

(d) Exactly as above, the $r-1$ edges $\{x_0, x_i^*\}$, $2 \leq i \leq r$, are in $E^r \setminus E$, implying that all the edges between x_0 and $W_0^* \cup W_1^*$ are in E, and that $|W_2^*| = \ldots = |W_r^*| = 1$. The proof of (d') is similar.

Thus, the six claims of Lemma H.9 are proved. □

We now distinguish between two cases, according to the type of G^*.

(i) G^* is of type 1 : $\delta^* = r+1$ and G^* is a sum $(W_0^*, W_1^*, \ldots, W_{r+1}^*)$ of cliques such that there are at most two values of i such that $|W_i^*| \geq 2$, and they are consecutive. Thanks to Lemma H.9(a), we can, without loss of generality, assume that x_0 is the neighbour of a vertex in W_0^* or in W_1^*, the case with W_r^*, W_{r+1}^* being similar.

If x_0 has no neighbour in W_1^*, then it has at least one in W_0^* and we can apply Lemma H.9(c) : all the edges between x_0 and W_0^* are in E, and $|W_1^*| = |W_2^*| = \ldots = |W_{r-1}^*| = 1$. If $|W_0^*| \geq 2$, then, because G^* is of type 1, we have : $|W_r^*| = |W_{r+1}^*| = 1$, and G, which has diameter $r+2$, is of type 2 with $W_0 = \{x_0\}$ and $W_i = W_{i-1}^*$, $1 \leq i \leq r+2$, see Figure III(a). If $|W_0^*| = 1$, then G, of diameter $r+2$, is of type 2 with $W_0 = \{x_0\}$ and $W_i = W_{i-1}^*$, $1 \leq i \leq r+2$, see Figure III(b).

If x_0 has a neighbour in W_1^*, we can apply Lemma H.9(d) : all the edges between x_0 and $W_0^* \cup W_1^*$ are in E, and $|W_2^*| = \ldots = |W_r^*| = 1$. If $|W_1^*| \geq 2$, then $|W_{r+1}^*| = 1$ because G^* is of type 1, and G, which has diameter $r+1$, is of type 1 with $W_0 = W_0^* \cup \{x_0\}$ and $W_i = W_i^*$, $1 \leq i \leq r+1$, see Figure IV(a). If $|W_1^*| = 1$, then G, which has diameter $r+1$,

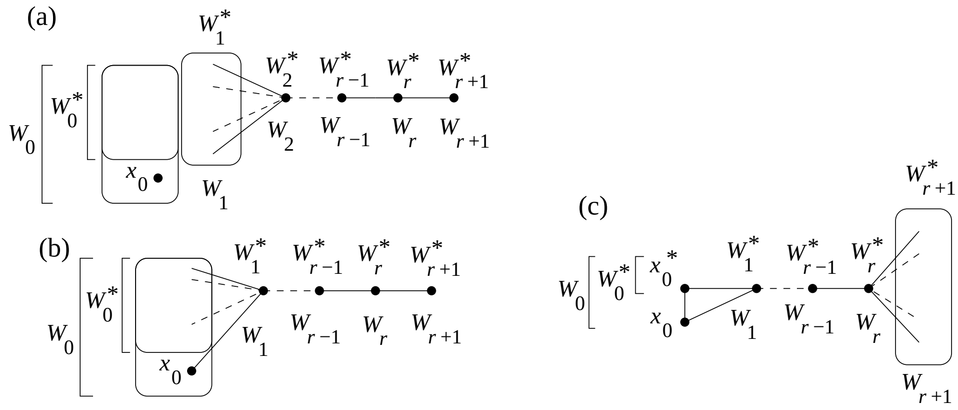


Figure IV – *G^* is of type 1 and x_0 has a neighbour in W_1^* ; (a) $|W_1^*| \geq 2$ and G is of type 1 ; (b)(c) $|W_1^*| = 1$ and G is of type 2 with $\delta = r+1$.*

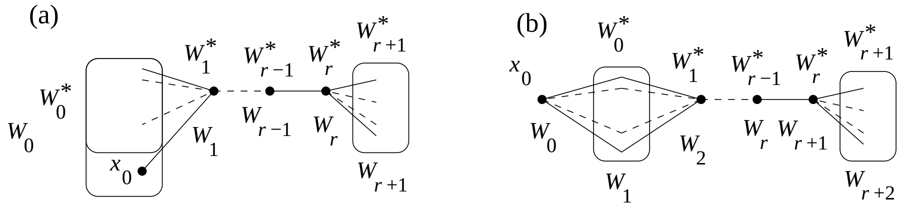


Figure V – *G^* is of type 2 with $\delta^* = r+1$; (a) $\{x_0, x_1^*\}$ belongs to E and G is of type 2 with $\delta = r+1$; (b) $\{x_0, x_1^*\}$ does not belong to E and G is of type 2 with $\delta = r+2$.*

is of type 2 with $W_0 = W_0^* \cup \{x_0\}$ and $W_i = W_i^*$, $1 \leq i \leq r+1$, see Figure IV(b)(c) : note that $|W_0^*| \geq 2$ and $|W_{r+1}^*| \geq 2$ cannot occur both, since G^* is of type 1.

This settles the case (i), when G^* is of type 1. When G^* is of type 2, there are three subcases, according to the diameter δ^* of G^*.

(ii) G^* is of type 2.
(A) $\delta^* = r+1$; then G^* is a sum $(W_0^*, W_1^*, \ldots, W_{r+1}^*)$ of cliques such that $|W_i^*| = 1$ for $1 \leq i \leq r$. Again, we can assume, without loss of generality, that x_0 is the neighbour of a vertex in W_0^* or in $W_1^* = \{x_1^*\}$.

If $\{x_0, x_1^*\} \in E$, then Lemma H.9(d) shows that every edge between x_0 and W_0^* belongs to E, and finally that G, which has diameter $r+1$, is of type 2 with $W_0 = W_0^* \cup \{x_0\}$ and $W_i = W_i^*$, $1 \leq i \leq r+1$, see Figure V(a).

If $\{x_0, x_1^*\} \notin E$, then by Lemma H.9(c), all the edges between x_0 and W_0^* are in E, and G, which has diameter $r+2$, is of type 2 with $W_0 = \{x_0\}$ and $W_i = W_{i-1}^*$, $1 \leq i \leq r+2$, see Figure V(b).

(B) $\delta^* = r+2$; then G^* is a sum $(W_0^*, W_1^*, \ldots, W_{r+2}^*)$ of cliques such that $|W_i^*| = 1$ for $2 \leq i \leq r+1$: without loss of generality, we can assume that it is W_{r+1}^* which is a singleton. Since here, unlike previously, 0 and 1 do not play the same role as $r+1$ and $r+2$, we have to consider whether x_0 has a neighbour in W_0^*, W_1^*, W_{r+1}^* or in W_{r+2}^*. In each case, Lemma H.9 is used.

If x_0 has a neighbour in W_0^* and not in W_1^*, then all the edges between x_0 and W_0^* are in E, $|W_1^*| = 1$, and G, which has diameter $r+3$, is of type 2 with $W_0 = \{x_0\}$ and $W_i = W_{i-1}^*$, $1 \leq i \leq r+3$, see Figure VI(a).

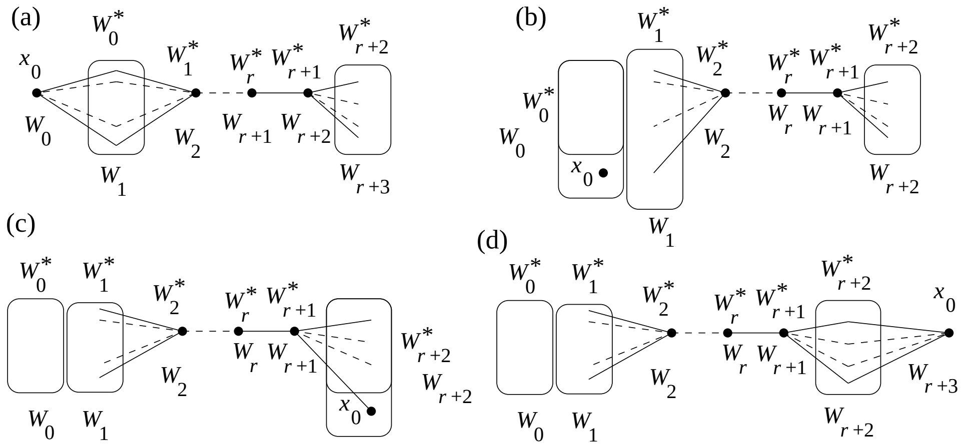


Figure VI – *G^* is of type 2 with $\delta^* = r+2$; (a) x_0 is neighbour to $W_0^* \setminus W_1^*$ and G is of type* 2 *with $\delta = r+3$; (b) x_0 is neighbour to W_1^* and G is of type* 2 *with $\delta = r+2$; (c) $\{x_0, x_{r+1}^*\} \in E$ and G is of type* 2 *with $\delta = r+2$; (d) x_0 is neighbour to $W_{r+2}^* \setminus \{x_{r+1}^*\}$ and G is of type* 2 *with $\delta = r+3$.*

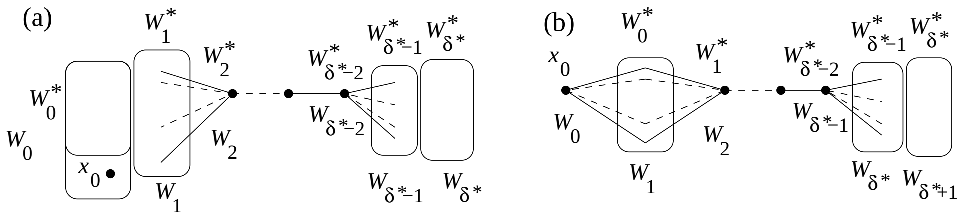


Figure VII – *G^* has type* 2 *with $\delta^* > r+2$; (a) x_0 is linked to W_1^*, G has type* 2 *with $\delta = \delta^*$; (b) x_0 is not linked to W_1^*, G has type* 2 *with $\delta = \delta^* + 1$.*

If x_0 has a neighbour in W_1^*, then G is of type 2 with diameter $r+2$, with $W_0 = W_0^* \cup \{x_0\}$ and $W_i = W_i^*$, $1 \leq i \leq r+2$, see Figure VI(b).

If x_0 has x_{r+1}^* for neighbour, then G is of type 2 with diameter $r+2$, with $W_i = W_i^*$, $0 \leq i \leq r+1$, and $W_{r+2} = W_{r+2}^* \cup \{x_0\}$, see Figure VI(c).

If x_0 has a neighbour in W_{r+2}^* and is not neighbour to x_{r+1}^*, then G is of type 2 with diameter $r+3$, with $W_i = W_i^*$, $0 \leq i \leq r+2$, and $W_{r+3} = \{x_0\}$, see Figure VI(d).

(C) $\delta^* > r+2$; then G^* is a sum $(W_0^*, W_1^*, \dots, W_{\delta^*}^*)$ of cliques with $|W_i^*| = 1$ for $2 \leq i \leq \delta^* - 2$. Here we can again assume, without loss of generality, that x_0 is the neighbour of a vertex in W_0^* or in W_1^*, and we use Lemma H.9.

If x_0 has a neighbour in W_1^*, then G has diameter δ^* and is of type 2, with $W_0 = W_0^* \cup \{x_0\}$ and $W_i = W_i^*$, $1 \leq i \leq \delta^*$, see Figure VII(a).

If x_0 has a neighbour in W_0^* and none in W_1^*, then necessarily $|W_1^*| = 1$, G has diameter $\delta^* + 1$ and is of type 2, with $W_0 = \{x_0\}$ and $W_i = W_{i-1}^*$, $1 \leq i \leq \delta^* + 1$, see Figure VII(b).

This ends the proof of Theorem H.8. □

Now that we have characterized the graphs reaching $b(r,n)$, we address question (III)(iii) from Section H.1.2, which was already raised, for $r = 2$, in [1]. The answer, given in the following theorem, is that any size is possible. In this theorem, the condition on n is due to the fact that $\text{diam}(G) \geq r+1$, since $G^r \neq K_n$; the condition $m \geq n-1$ comes from

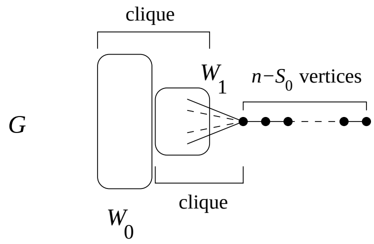


Figure VIII – *The graph G constructed for Theorem H.10.*

the connectivity of the graphs considered; and, also because $G^r \neq K_n$, we must have $m + b(r,n) \leq n(n-1)/2 - 1$, which gives the second inequality in (H.7).

Theorem H.10. *Let r, n, m be such that $n \geq r+2$ and*

$$n - 1 \leq m \leq \frac{(n-r+1)(n-r)}{2} + r - 2. \qquad \text{(H.7)}$$

There exists a graph G of order n and size m such that $|E^r \setminus E| = b(r,n)$.

Proof. We consider the function

$$f \;:\; S \longrightarrow \frac{(S-1)(S-2)}{2} + n - 1, \quad \text{for } S \in \{2, 3, \ldots, n-r+1\}.$$

This function is strictly increasing and verifies

$$f(2) = n - 1 \;\text{ and }\; f(n-r+1) = \frac{(n-r+1)(n-r)}{2} + r - 1;$$

therefore, by (H.7), we have : $f(2) \leq m < f(n-r+1)$. Given m, there is a unique S between 2 and $n-r$ such that $f(S) \leq m < f(S+1)$; let S_0 be this value :

$$\frac{(S_0-1)(S_0-2)}{2} + n - 1 \leq m \leq \frac{S_0(S_0-1)}{2} + n - 2. \qquad \text{(H.8)}$$

Let $\alpha = \frac{S_0(S_0-1)}{2} + n - 1 - m$ and $\beta = S_0 - \alpha$. Using the first inequality in (H.8), we see that $\alpha \leq S_0 - 1$, so $\beta \geq 1$; the second inequality in (H.8) yields $\alpha \geq 1$. We are now ready to give a graph with n vertices and m edges which meets the bound $b(r,n)$: consider the graph G of order n given by Figure VIII, with $|W_0| = \alpha \geq 1$, $|W_1| = \beta \geq 1$, and $\alpha+\beta = S_0 \in \{2, 3, \ldots, n-r\}$. The size of G is $\frac{S_0(S_0-1)}{2} + \beta + (n - S_0 - 1) = \frac{S_0(S_0-1)}{2} - \alpha + n - 1$, which is equal to m, using the definition of α. The diameter of G is $n - S_0 + 1$, which is at least $r+1$ since S_0 is at most $n-r$, and therefore $G^r \neq K_n$. Finally, either by straightforwardly counting the edges in G^r, or by observing that G belongs to one of the two types of graphs given in Theorem H.8 and illustrated in Figures I and II (with $i = 0$ in Figure I, $|W_{\delta-1}| = |W_\delta| = 1$ in Figure II(a), or $|W_{r+2}| = 1$ in Figure II(b), according to the diameter of G), one can see that G indeed meets the bound $b(r,n)$. □

Looking back at the proof of Theorem H.10, we observe that, given m, we first determined S_0, the cardinality of $W_0 \cup W_1$, and then we distributed the S_0 vertices among W_0 and W_1, α vertices in W_0 and β vertices in W_1, in order to obtain the desired number of edges.

H.3 Conclusion

We have addressed three problems on the sizes of G and G^r.

In Section H.1.3 we have given the known results on the size of a graph with a given diameter.

In Section H.2.1 we have characterized the graphs which attain the lower bound (H.1) on the size of G^r.

Finally, in Section H.2.2, we have answered the following question : given an integer $r \geq 2$ and a connected graph G of order n such that $G^r \neq K_n$, what is the minimum number of edges that are added when going from G to G^r ? We have also characterized the graphs achieving this bound and proved that they can reach all possible sizes.

Annexe I

On the Sizes of G, G^r, $G^r \setminus G$, Part II : the Directed Case

David Auger [1], **Irène Charon** [1],
Olivier Hudry [1], **Antoine Lobstein** [2]
{david.auger, irene.charon, olivier.hudry, antoine.lobstein}@telecom-paristech.fr

Abstract

Let G be a directed graph and G^r be its r-th power. We study different issues dealing with the number of arcs, or size, of G and G^r : given the order and diameter of a strongly connected digraph, what is its maximum size, and which are the graphs achieving this bound ? what is the minimum size of the r-th power of a strongly connected digraph, and which are the graphs achieving this bound ? given a strongly connected digraph G of order n such that $G^r \neq K_n$, what is the minimum number of arcs that are added when going from G to G^r, and which are the graphs achieving this bound ?

Keywords : raph Theory, Directed Graph, Digraph, Diameter, Power of a Graph.

I.1 Introduction

We first give some very basic definitions and notation for directed graphs, before we expound our study.

I.1.1 Definitions and Notation

We shall denote by $G = (V, A)$ a *directed graph*, or *digraph*, with vertex set V and arc set A, where an *arc* from $x \in V$ to $y \in V$ is denoted by the couple (x, y) ; we say that x is the *origin* of the arc, and y its *end*. We require the graph to have no loops nor double arcs

2. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

— but the arcs (x,y) and (y,x) can of course simultaneously exist. The *size* of a digraph is its number of arcs, and its *order* is its number of vertices.

A directed *path* $P = x_0x_1 \dots x_\ell$ is a sequence of vertices x_i, $0 \le i \le \ell$, such that $(x_i, x_{i+1}) \in A$ for i between 0 and $\ell - 1$. The *length* of P is its number of arcs, ℓ. A digraph is called *strongly connected* if for any two vertices x and y there is a path going from x to y.

In a strongly connected graph G, we can define the distance from any vertex x to any vertex y, denoted by $d_G(x,y)$, as the number of arcs in any shortest directed path from x to y, since such a path always exists. Note that in general, $d_G(x,y)$ is not equal to $d_G(y,x)$. The *diameter* of a strongly connected graph G is the maximum distance in the graph :

$$\mathrm{diam}(G) = \max_{x \in V, y \in V} d_G(x,y).$$

Given an integer $r \ge 1$, the *r-th power*, or *r-th transitive closure*, of the graph $G = (V,A)$ is the graph $G^r = (V, A^r)$, where, for two distinct vertices x and y, the arc (x,y) is in A^r if and only if $d_G(x,y) \le r$.

The *clique*, or *complete graph*, K_n, is the digraph of order n with all possible $n(n-1)$ arcs. Finally, an *induced subgraph* of $G = (V,A)$ is a graph $G^* = (V^*, A^*)$ where $V^* \subseteq V$ and $A^* = \{(x,y) : x \in V^*, y \in V^*, (x,y) \in A\}$; a *subgraph* is such that A^* is included in $\{(x,y) : x \in V^*, y \in V^*, (x,y) \in A\}$.

I.1.2 Scope of the Paper

We are interested in the following related problems on sizes and powers, for digraphs :

(a) Given the order and diameter of a strongly connected digraph, what is its maximum size, and which are the graphs achieving this bound ? To our knowledge, this very natural problem has not been studied, and we give the answer in Section I.2.

(b) Given an integer $r \ge 2$, what is the minimum size of a digraph of order n, of which it is known that it is the r-th power of a strongly connected digraph, and which are the graphs achieving this bound ? The complete answer is in Section I.3.

(c) Given an integer $r \ge 2$ and a strongly connected digraph G of order n such that $G^r \neq K_n$, what is the minimum number of arcs that are added when going from G to G^r, and which are the graphs achieving this bound ? We give the answer in Section I.4.

The same issues for undirected graphs are treated in [11].

I.2 The Size of a Digraph with Given Order and Diameter

Unexpectedly, the following result, although easy, is new, as far as we know.

Theorem I.1. *Let $G = (V,A)$ be a strongly connected digraph of order n and diameter $\delta \ge 2$. Then the size of G is at most*

$$\frac{\delta(\delta+3)}{2} + (n - \delta - 1)(n+2).$$

Proof. Let $z_1, z_2 \in V$ be such that $d_G(z_1, z_2) = \delta$, and C be a shortest directed path from z_1 to z_2 : $C = x_0x_1 \dots x_\delta$, with $x_0 = z_1$ and $x_\delta = z_2$; there are no more arcs (x_i, x_j), $i < j$, but all the arcs (x_i, x_j), $i > j$, can exist. In G, the remaining vertices y_j, $1 \le j \le n - \delta - 1$, can at most constitute the clique $K_{n-\delta-1}$, and each y_j can be part of at most $\delta + 4$ arcs with ends or origins in C : this is clear if $\delta = 2$; if $\delta \ge 3$ and there are $\delta + 5$ arcs (y_j, x_k) or

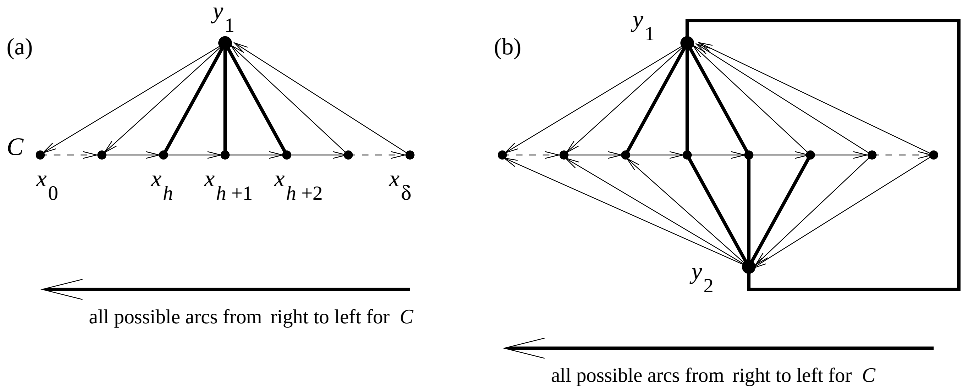


Figure I – *The location of the arcs between the vertices in C and the clique. A bold line without arrow represents a double arc.*

(x_k, y_j), then there are at least four vertices x_i such that both (y_j, x_i) and (x_i, y_j) belong to A. This in turn implies that in A there are two arcs (x_{i_1}, y_j) and (y_j, x_{i_2}) with $i_1+3 \leq i_2$. This is impossible, since the path $x_0 \dots x_{i_1} y_j x_{i_2} \dots x_\delta$ would be shorter than C. All in all, we have at most

$$\delta + \frac{\delta(\delta+1)}{2} + (n-\delta-1)(n-\delta-2) + (\delta+4)(n-\delta-1)$$

arcs in G. □

We shall see that this theorem is also a direct consequence of Theorem I.4. We set

$$\sigma(\delta, n) = \frac{\delta(\delta+3)}{2} + (n-\delta-1)(n+2), \tag{I.1}$$

and we are going to characterize the graphs $G = (V, A)$ reaching $\sigma(\delta, n)$. The previous proof shows that necessarily G consists of the path $C = x_0 x_1 \dots x_\delta$, all the arcs (x_i, x_j), $i > j$, the clique $K_{n-\delta-1}$ and exactly $\delta + 4$ arcs between every vertex $y \in K_{n-\delta-1}$ and the vertices of C. All we have to determine is how these $(\delta+4)(n-\delta-1)$ arcs are located.

We observe that, in particular, there are in A at least three arcs (x_i, y) and three arcs (y, x_j). Let h be the smallest subscript such that there is a vertex $y_1 \in K_{n-\delta-1}$ with $(x_h, y_1) \in A$; the parameter h can vary from 0 to $\delta - 2$. Let k be the largest subscript such that $(y_1, x_k) \in A$. Because $x_h y_1 x_k$ must not allow shortcuts with respect to C, we have $k \leq h+2$. Now one can see that everything is forced : $k = h+2$, and we must have in A all the $\delta - h + 1$ arcs (x_i, y_1), $h \leq i \leq \delta$, and all the $h+3$ arcs (y_1, x_j), $0 \leq j \leq h+2$, see Figure I(a).

Now consider another vertex y_2 in the clique. If $(x_h, y_2) \in A$, then everything goes as with y_1. If $(x_h, y_2) \notin A$, let ℓ (respectively, m) be the largest (respectively, smallest) subscript such that $(y_2, x_\ell) \in A$ (respectively, $(x_m, y_2) \in A$). Because of the forbidden shortcut $x_h y_1 y_2 x_\ell$, we have $\ell \leq h+3$. And again, there is no choice left : $\ell = h+3$, $m = \ell - 2 = h+1$, and we must have in A all the $h+4$ arcs (y_2, x_j), $0 \leq j \leq h+3$, and all the $\delta - h$ arcs (x_i, y_2), $h+1 \leq i \leq \delta$, see Figure I(b).

So the clique is divided into at most two types of vertices, those with $(x_h, y) \in A$, and the others, with $(x_h, y) \notin A$ but $(x_{h+1}, y) \in A$. When h varies from 0 to $\delta - 2$, we obtain the description of all the graphs achieving the bound $\sigma(\delta, n)$. Note that if $h = \delta - 2$, we can have only one type of vertices in the clique.

I.3 Size of the Power of a Digraph

We address the following issue : given a strongly connected digraph of order n, $G = (V, A)$, what is the smallest number of arcs in G^r ? and which are the graphs which meet this bound ?

We give the complete answer in the next two theorems. If $r \geq n-1$, then $G^r = K_n$ and the problem is trivial, so we assume that $r \leq n-2$.

Theorem I.2. *If $r \leq n-2$ and $G = (V, A)$ is a strongly connected digraph of order n, then the size of G^r is at least nr.*

Proof. Let $x \in V$. If for all $y \in V$, $d_G(x, y) \leq r$, then the $n-1$ arcs (x, y), $y \in V \setminus \{x\}$, are in G^r. If there is a vertex y such that $d_G(x, y) > r$, consider a shortest path $xz_1z_2 \ldots z_r \ldots y$ from x to y. Then the r arcs (x, z_i), $1 \leq i \leq r$, are in G^r. In both cases, we see that a vertex x brings at least r arcs to G^r, since $r \leq n-1$. □

Theorem I.3. *If $r \leq n-2$, the only strongly connected digraph $G = (V, A)$ with order n such that G^r has size exactly nr is the* circuit $x_0x_1 \ldots x_{n-1}x_0$.

Proof. Obviously, the circuit achieves rn. If G meets rn, then each vertex x in G must contribute exactly r to G^r. Since $n-1 > r$, we are in the case when there is a shortest path $xz_1z_2 \ldots z_r \ldots y$ from x to some vertex $y \in V$. If there is an arc $(x, w) \in A$, $w \neq z_1$, then x gives at least $r+1$ arcs to G^r. So in A the only arc with origin x is (x, z_1) and we have just proved that every vertex is the origin of exactly one arc in G. Since G is strongly connected, the only possibility is the circuit. □

I.4 From G to G^r

We consider an integer $r \geq 2$ and all strongly connected digraphs $G = (V, A)$, of order n, such that $G^r \neq K_n$, and we want to determine what is the minimum number of arcs that have to be added to go from G to G^r, i.e., what is the minimum cardinality of $A^r \setminus A$: we shall denote this number by $\Lambda(r, n)$. Once we know $\Lambda(r, n)$, we shall characterize the graphs reaching it.

Observe that the condition $G^r \neq K_n$ implies that G has diameter at least $r+1$. In the following theorem, we can see that $\Lambda(r, n)$ is linear in n, with the factor $r-1$.

Theorem I.4.

$$\Lambda(r, n) = (r-1)(n-1-\frac{r}{2}). \qquad \text{(I.2)}$$

Proof. Note that equality (I.2) contains the case $r = 1$. We set

$$b(r, n) = (r-1)(n-1-\frac{r}{2}).$$

First, we exhibit a digraph G_0 showing that $\Lambda(r, n) \leq b(r, n)$. This graph has vertex set $V_0 = \{x_i : 0 \leq i \leq n_0 - 1\}$ and arc set

$$A_0 = \{(x_i, x_{i+1}) : 0 \leq i \leq n_0 - 2\} \cup \{(x_j, x_k) : 0 \leq k < j \leq n_0 - 1\}. \qquad \text{(I.3)}$$

In other words, G_0 consists of a directed path going from x_0 to x_{n_0-1}, plus all arcs going from a vertex x_j to a vertex of smaller subscript.

Because G_0 has diameter at least $r+1$, we have $n_0 \geq r+2$. The arcs in $A_0^r \setminus A_0$ are the arcs (x_i, x_j) with $0 \leq i \leq n_0 - 3$ and $i+2 \leq j \leq \min\{i+r, n_0-1\}$. So there are $r-1$

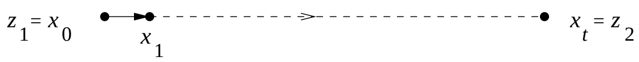


Figure II – *The path C used in the proof of Theorem I.4.*

additional arcs starting from x_i as long as $i \leq n_0 - 1 - r$, and $(r-2), (r-3), \ldots, 1$ additional arcs starting from the subsequent vertices, $x_{n_0-r}, x_{n_0-r+1}, \ldots, x_{n_0-3}$, respectively. All in all, we have $(n_0 - r)(r-1) + \frac{1}{2}(r-1)(r-2) = b(r, n_0)$ new arcs, which proves the upper bound for $\Lambda(r, n)$.

Now, let $G = (V, A)$ be any digraph fulfilling the hypotheses, and $G^* = (V^*, A^*)$ be a strongly connected induced subgraph of G, having minimum order, n^*, among all the strongly connected induced subgraphs of G which have two vertices at distance in G greater than r from one another — since G has diameter at least $r+1$, such two vertices exist and if necessary we take $G^* = G$, so a graph G^* always exists. We name z_1 and z_2 these two vertices, so that $z_1 \in V^*$, $z_2 \in V^*$ and $d_G(z_1, z_2) > r$. Obviously, $d_{G^*}(z_1, z_2) \geq d_G(z_1, z_2) > r$.

It is useful to give a name to the following property, which simply uses the very definition of G^* :

> (P) If H is a strongly connected subgraph of G^* with order n_H such that $2 \leq n_H < n^*$, then any two vertices x, y in H are at distance in G at most r from one another, and $(x, y) \in A^r$.

In a first step, we are going to show that V^* contains at least $b(r, n^*)$ couples of vertices u, v such that the arc (u, v) belongs to $A^r \setminus A$. These couples will be called *friendly couples*.

Let C be a shortest path in G^* from z_1 to z_2 :

$$C = x_0 x_1 \ldots x_t, \tag{I.4}$$

with $x_0 = z_1, x_t = z_2$ and $t > r$, see Figure II.

The path C is a subpath of G_0 with $n_0 - 1 = t$: same vertices, fewer arcs ; therefore, there are at least the $b(r, t+1)$ arcs (x_i, x_j), $i + 2 \leq j$, to be added when going from G^* to $(G^*)^r$, *cf.* the above study of G_0. Moreover, these couples x_i, x_j are friendly : (a) $x_i, x_j \in V^*$, (b) $(x_i, x_j) \notin A$ because $(x_i, x_j) \notin A^*$, and (c) $(x_i, x_j) \in A^r$ because $(x_i, x_j) \in (A^*)^r \subseteq A^r$. So :

$$\text{there are at least } b(r, t+1) \text{ friendly couples } x_i, x_j \text{ in } C. \tag{I.5}$$

If any vertex in G^* belongs to C, i.e., $n^* = t + 1$, then we have proved the existence of at least $b(r, n^*)$ friendly couples in G^*. So from now on, we can assume that there is at least one vertex in G^* which is not a vertex $x_i, 0 \leq i \leq t$. We denote by Y^* the set of these vertices :

$$Y^* = V^* \setminus \{x_i : 0 \leq i \leq t\}.$$

In an intermediate step, our goal is to prove the following property (Q) :

> (Q) for any vertex $y \in Y^*$, we add, when going from G to G^r, at least $r - 1$ distinct arcs such that :
>
> (i) either the origin is y and the end is in G^*,
>
> (ii) or the end is y and the origin is in C.

As a consequence, we never count twice the same arc for two different y_1, y_2 in Y^*. Obviously, these arcs yield friendly couples. To prove that (Q) is true, we state a first lemma.

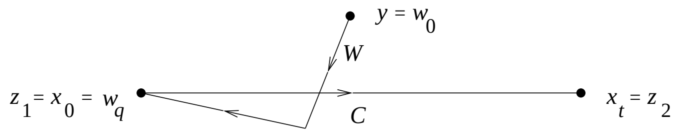


Figure III – *The paths C and W used in the proof of Theorem I.4.*

Lemma I.5. *Let $y \in Y^*$. If for all x_i in C, $d_G(x_i, y) \leq r$ and $d_G(y, x_i) \leq r$, then property (Q) holds.*

Proof of Lemma I.5. By assumption, all the arcs (x_i, y) and (y, x_i) belong to A^r, so all we have to show is that at least $r - 1$ of them do not belong to A.

If no i exists such that $(x_i, y) \in A$, then (Q) holds, because $t > r$. So we can assume that there is a smallest k, $0 \leq k \leq t$, such that $(x_k, y) \in A$. We use the fact, which is true for all j between $k+3$ and t, that $(y, x_j) \notin A$: otherwise the two arcs $(x_k, y), (y, x_j)$, belonging to $A^* \subseteq A$, would contribute to provide, in G^*, a path shorter than C from z_1 to z_2.

Therefore, the arcs (y, x_j), $k + 3 \leq j \leq t$, and (x_i, y), $0 \leq i \leq k - 1$, do not belong to A, i.e., all in all, $t - 2 \geq r - 1$ arcs, which proves Lemma I.5. □

Now we consider a shortest path W in G^* from y to x_0 :

$$W = w_0 w_1 \dots w_q,$$

with $w_0 = y$ and $w_q = x_0$, see Figure III. Note that the intersection between C and W is not necessarily reduced to x_0. For i between 2 and q, the arc (y, w_i) does not belong to A, because W is a shortest path, and for i between 2 and $\min\{q, r\}$, the arc (y, w_i) belongs to A^r and satisfies (i). If $q \geq r$, then (Q) holds, so from now on we assume that $q < r$, and we have just shown that

$$\text{in } A^r \setminus A, \text{ there are } q - 1 \text{ arcs } (y, w_i),\ 2 \leq i \leq q, \text{ satisfying (i).} \tag{I.6}$$

We consider the vertices $x_0, x_1, \dots, x_{r-q}$ in C. By triangle inequality, for i between 0 and $r - q$, we have $d_{G^*}(y, x_i) \leq q + (r - q) = r$ and so

$$d_G(y, x_i) \leq r \text{ for } 0 \leq i \leq r - q. \tag{I.7}$$

We are now ready to prove Lemma I.6.

Lemma I.6. *If there exists a path in G^* from y to x_k, $1 \leq k \leq r - q$, which does not go through x_0, then (Q) is true.*

Proof of Lemma I.6. Among these paths, we choose the path U for which k is minimum, see Figure IV. For i between 1 and $k - 1$, the $k - 1$ arcs (y, x_i) are in $A^r \setminus A$, because of (I.7), and satisfy (i), so we have just shown that

$$\text{there are } k - 1 \text{ arcs } (y, x_i) \text{ in } A^r \setminus A \text{ satisfying (i).} \tag{I.8}$$

Let M be a path in G^* from x_t to y, see Figure V. Then M does not go through x_0 : otherwise, consider the subpath M^* of M going from x_t to x_0, and $C \cup M^*$, by which we mean the induced subgraph of G with vertices in $C \cup M^*$, so that there can be more arcs than simply the arcs of the path C and the arcs of the subpath M^*. Now this graph is

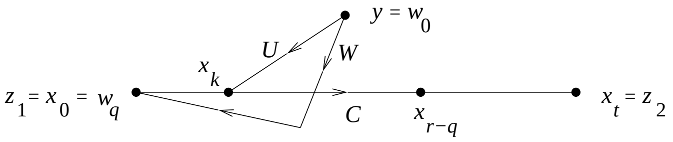


Figure IV – *The paths C, W and U used in the proof of Theorem I.4.*

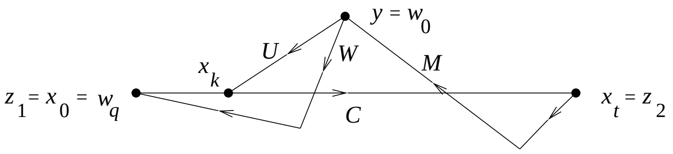


Figure V – *The paths C, W, U and M used in the proof of Theorem I.4.*

strongly connected, contains $x_0 = z_1$ and $x_t = z_2$ which are at distance greater than r in G, and does not contain y, thus contradicting the minimality of G^*.

Consider next the subpath C^* of C going from x_k to x_t, see Figure VI, and $U \cup C^* \cup M$; this graph, which has, among others, arcs going from y to x_k, $k \neq 0$, from x_k to x_t, and from x_t to y, is strongly connected, contains fewer vertices than G^* (because it does not contain x_0), and so, by property (P), we have

$$d_G(y, x_i) \leq r \text{ and } d_G(x_i, y) \leq r, \quad k \leq i \leq t. \tag{I.9}$$

If for all i between $k+1$ and $t-1$, the arcs (x_i, y) are not in A, then they are in $A^r \setminus A$ and satisfy (ii). So, together with our first $k-1$ arcs from (I.8), we have at least $k-1+(t-k-1) = t-2 \geq r-1$ suitable arcs. We assume finally that there is an arc (x_h, y) in A, with $k+1 \leq h \leq t-1$, see Figure VII.

Consider, with abuse of notation, the graph $\mathcal{G} = (x_h, y) \cup W \cup \{x_0, x_1, \ldots, x_h\}$, which has, among others, arcs going from x_h to y, from y to x_0 and from x_0 back to x_h. This strongly connected graph cannot go through x_t, otherwise it would do so along W, and, since W goes back to x_0, we would again have the vertices in C strongly connected between themselves, contradicting the minimality of G^*. So this graph $\mathcal{G}$ has fewer vertices than G^*, and therefore by property (P), for all i between 0 and h, we have : $d_G(y, x_i) \leq r$, $d_G(x_i, y) \leq r$. Using (I.9) and $h \geq k+1$, we see that we are in the conditions of Lemma I.5, which shows that (Q) is true and ends the proof of Lemma I.6. □

Finally, we assume that the hypothesis of Lemma I.6 is not fulfilled : in particular, in A^*, hence in A, there is no arc (y, x_i) for i between 1 and $r-q$, but, because of (I.7), all these arcs are in A^r. So :

$$\text{in } A^r \setminus A, \text{ there are } r-q \text{ arcs } (y, x_i),\ 1 \leq i \leq r-q, \text{ satisfying (i).} \tag{I.10}$$

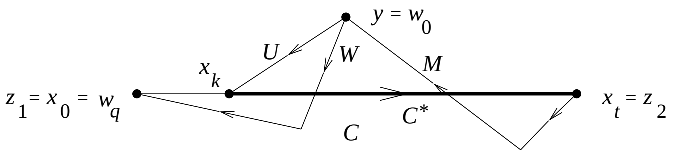


Figure VI – *The paths C, W, U, M and C^* used in the proof of Theorem I.4.*

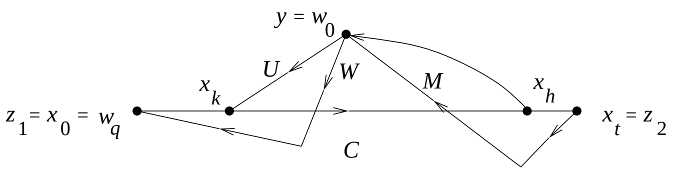


Figure VII – *The paths C, W, U, M, and the arc (x_h, y) used in the proof of Theorem I.4.*

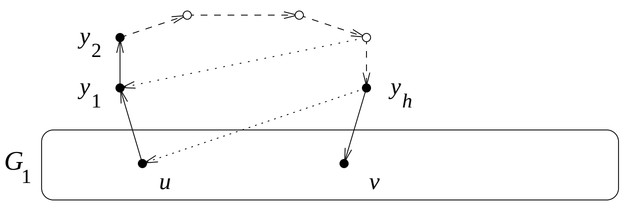


Figure VIII – *The graph G_1 and a smallest arch induce the graph G_2.*

Also because we are not in the conditions of Lemma I.6, and unlike in Figure IV, a vertex w_i in W, $2 \le i \le q$, cannot coincide with a vertex x_j, $1 \le j \le r-q$, for otherwise the beginning of the path W from y to x_0 would be a path from y to x_j not going through x_0. Therefore we can add the arcs obtained in (I.6) and (I.10), which proves that property (Q) holds in all cases.

Consequently, by (I.5) and because (Q) provides friendly couples which are counted only once, we have in G^* at least

$$b(r, t+1) + (r-1)\big(n^* - (t+1)\big) = b(r, n^*) \tag{I.11}$$

friendly couples, which ends our first step.

Now we consider all the strongly connected induced subgraphs of G, of order η, containing two vertices at distance in G greater than r from one another, and containing at least $b(r, \eta)$ friendly couples — we have just proved that such graphs exist ; among them, we take one, $G_1 = (V_1, A_1)$, with largest order, n_1. If $n_1 = n$, then $G_1 = G$, there are at least $b(r, n)$ arcs in $A^r \setminus A$, and Theorem I.4 is proved. So from now on, we assume that $V \setminus V_1 \neq \emptyset$. The resulting contradiction will prove Theorem I.4.

Because G is strongly connected, there is a smallest set of vertices $Y = \{y_1, y_2, \ldots, y_h\} \subseteq V \setminus V_1$ such that $H = y_1 y_2 \ldots y_h$ is a directed path in G and the arcs (u, y_1), (y_h, v) are in A, where u and v are two, non necessarily distinct, vertices in V_1 ; see Figure VIII. We call Y an *arch* and set $G_2 = (V_2, A_2)$, the induced subgraph of G with vertex set $V_2 = V_1 \cup Y$; the arc set A_2 contains, among others, the arcs of A_1 and of H, as well as (u, y_1) and (y_h, v). Because of the minimality of Y, there is in A_2, hence in A, no arc (y_i, y_j) with $1 \le i < j \le h$, no arc with origin in V_1 and end y_i, $2 \le i \le h$, and no arc with origin y_i, $1 \le i \le h-1$, and end in V_1.

We are going to show that the vertices in Y each give $r-1$ arcs to $A^r \setminus A$, except possibly one vertex ; in this case however, this will be compensated by one vertex giving $2(r-1)$ arcs. In any case, no arc will be counted twice.

We first assume that $h \ge 2$.

Let $y \in Y$. If there is a vertex x in G_2 such that $d_{G_2}(y, x) \ge r+1$, then

$$\text{in } A^r \setminus A, \text{ there are } r-1 \text{ arcs with origin } y. \tag{I.12}$$

Indeed, a shortest path in G_2 from y to x : $z_0 z_1 \ldots z_t$, with $z_0 = y$, $z_t = x$ and $t > r$, shows that the $r-1$ arcs (y, z_i), $2 \leq i \leq r$, belong to $A^r \setminus A$.

Let $y \in Y \setminus \{y_h\}$. If all vertices x in G_2 are such that $d_{G_2}(y,x) \leq r$, then

$$\text{in } A^r \setminus A, \text{ there are } r-1 \text{ arcs with origin } y. \tag{I.13}$$

Indeed, for all $x \in V_1$, the arc $(y,x) \notin A$, as mentioned above, and $(y,x) \in A^r$ because $d_{G_2}(y,x)$, hence $d_G(y,x)$, is at most r. Since G_1 contains two vertices at distance in G greater than r from one another, its order is at least $r-1$ and claim (I.13) is true.

Gathering (I.12) and (I.13), we obtain immediately the following lemma.

Lemma I.7. *If $h \geq 2$, then, for all $y \in Y \setminus \{y_h\}$,*

$$\textit{in } A^r \setminus A, \textit{ there are } r-1 \textit{ arcs with origin } y. \tag{I.14}$$

□

Lemma I.8. *If $h \geq 2$, then*

$$\textit{there are } r-1 \textit{ arcs in } A^r \setminus A \textit{ with origin in } V_1 \textit{ and end } y_2. \tag{I.15}$$

Proof of Lemma I.8. Take any vertex $x \in V_1$ and a shortest path C in G_2 from x to y_2. By the minimality of the arch Y, this path goes through vertices in V_1, then goes to y_1, and then it goes to y_2, by minimality of C.

First, we assume that there is a vertex $w \in V_1$ such that $d_{G_2}(w, y_2) \geq r$. By the above remark and because $r \geq 2$, there is a vertex $z \in V_1$ with $d_{G_2}(z, y_2) = r$ — and so $d_G(z, y_2) \leq r$ — and the first $r-1$ vertices in a shortest path in G_2 from z to y_2 belong to V_1. If we call these $r-1$ vertices z_0, z_1, …, z_{r-2}, then the arcs (z_i, y_2) belong to $A^r \setminus A$, and claim (I.15) holds.

If, on the other hand, for all $w \in V_1$, $d_{G_2}(w, y_2) < r$, then for all vertices w in V_1, $(w, y_2) \in A^r \setminus A$, and (I.15) follows, which proves Lemma I.8. □

Corollary I.9. *If $h \geq 2$, then there are at least $|Y|(r-1)$ distinct arcs in $A^r \setminus A$ with one end or one origin in Y.*

Proof of Corollary I.9. Simply add up the arcs obtained in (I.14) and in (I.15) : if $h \geq 3$, then y_h gives no arc, y_2 gives $r-1$ arcs with origin y_2 and $r-1$ arcs with origin in V_1 and end y_2, and the remaining vertices $y_i \in Y$ each give $r-1$ arcs with origin y_i ; if $h = 2$, then y_1 gives $r-1$ arcs with origin y_1, and y_2 gives $r-1$ arcs with origin in V_1 and end y_2. All these arcs are distinct, which proves Corollary I.9. □

We are left with the case $h = 1$.

Lemma I.10. *If $Y = \{y\}$, then there are, in $A^r \setminus A$, $r-1$ arcs whose origin or end is y.*

Proof of Lemma I.10. Assume first that there is a vertex x in G_1 such that $d_{G_2}(y,x) \geq r$; then the argument leading to (I.12) still works, and we obtain $r-1$ arcs in $A^r \setminus A$ with origin y. Similarly, if there is a vertex x in G_1 such that $d_{G_2}(x,y) \geq r$, then there exist $r-1$ arcs in $A^r \setminus A$ with end y.

Finally, we treat the case when for all $x \in V_1$, $d_{G_2}(x,y) < r$ and $d_{G_2}(y,x) < r$. We know that in V_1 there are two vertices at distance in G at least $r+1$ from one another : if we denote them by z_1 and z_2, there is in G_1 a shortest path $x_0 x_1 \ldots x_t$ with $x_0 = z_1$, $x_t = z_2$ and $t > r$. Mimicking the proof of Lemma I.5, we see that y is the origin or the end of $r-1$ arcs (y, x_i) or (x_i, y), which proves Lemma I.10. □

Corollary I.11. *For all $h \geq 1$, there are at least $h(r-1)$ distinct arcs in $A^r \setminus A$ with one end or one origin in the arch $Y = \{y_1, y_2, \dots, y_h\}$.* □

So in G_2, which has $n_1 + h$ vertices, there are two vertices at distance in G at least $r+1$ from one another, and there are $b(r, n_1) + h(r-1) = b(r, n_1 + h)$ arcs in $A^r \setminus A$, which contradicts the maximality of G_1 and ends the proof of Theorem I.4. □

Theorem I.4 implies directly Theorem I.1 : let $G = (V, A)$ be a strongly connected digraph of order n and diameter $\delta \geq 2$. Then $G^{\delta-1} \neq K_n$, $|A^{\delta-1}| \leq n(n-1) - 1$, and $|A| \leq |A^{\delta-1}| - b(\delta-1, n)$. Calculations show that $n(n-1) - 1 - b(\delta-1, n) = \sigma(\delta, n)$, including the case $\delta = 2$.

We now characterize the graphs which attain the bound $|A^r \setminus A| = \Lambda(r, n)$; we have already seen at the beginning of the proof of Theorem I.4 that the graph G_0 with vertex set $V_0 = \{x_i : 0 \leq i \leq n_0 - 1\}$ and arc set A_0 given by (I.3) is such that there are exactly $\Lambda(r, n_0)$ arcs in $A_0^r \setminus A_0$.

We consider a strongly connected digraph $G = (V, A)$ of order n, such that $G^r \neq K_n$ and $|A^r \setminus A| = \Lambda(r, n)$; the diameter δ of G is at least $r+1$.

In the process of proving Theorem I.4, we considered in (I.4) a shortest directed path $C = x_0 x_1 \dots x_t$, $t > r$, from $z_1 = x_0$ to $x_t = z_2$; this path will provide at least $\Lambda(r, t+1)$ friendly couples, cf. (I.5). Each vertex in $Y^* = V^* \setminus \{x_i : 0 \leq i \leq t\}$ will bring at least $r-1$ friendly couples, thanks to property (Q), and all in all G^* will give at least $\Lambda(r, n^*)$ friendly couples, cf. (I.11). Then, switching from G^* to G_1, we proved (Corollary I.11) that each vertex in $V \setminus V_1$ gives, in average, at least $r-1$ arcs to $A^r \setminus A$, finally leading to at least $\Lambda(r, n)$ arcs in $A^r \setminus A$.

If G attains the bound, then in the previous paragraph, we can replace each occurrence of “at least” by “exactly”. In particular, G^* achieves the bound $\Lambda(r, n^*)$ and C achieves the bound $\Lambda(r, t+1)$, for the number of friendly couples. Also, we can see that if $i > j$ and $(x_i, x_j) \in A^r$, then $(x_i, x_j) \in A$, because the $\Lambda(r, t+1)$ friendly couples are of type x_k, x_ℓ, $k < \ell$. We now show that all vertices of G^* are in C.

Lemma I.12. *If G achieves the bound $\Lambda(r, n)$, then the set Y^* is empty.*

Proof. Let D be a shortest path in G^* going from x_t to x_0. Then all the vertices in Y^* are vertices of D, for otherwise the set of vertices of C and D would violate the minimality of G^*. We now show that D has no vertices outside of C.

Assume the contrary.

If D intersects C in a vertex x_h, $h \notin \{0, t\}$, we consider the couple x_t, x_h. The arc (x_t, x_h) belongs to A^r : the graph consisting of $x_h x_{h+1} \dots x_t$ and the part of D from x_t to x_h is strongly connected and smaller than G^*, and we use property (P). Therefore, $(x_t, x_h) \in A$. Similarly, $(x_h, x_0) \in A$, and, because $r \geq 2$, (x_t, x_0) is in A^r, hence in A, which yields a path shorter than D.

Assume now that D does not intersect C.

If D, apart from x_t and x_0, has at most $r-1$ vertices, then the distance in D from x_t to x_0 is at most r, from which we can conclude that $(x_t, x_0) \in A$, again a contradiction.

So we assume that D, apart from x_t and x_0, has at least r vertices, $w_1, w_2, \dots, w_r, \dots$, with (x_t, w_1) the first arc in D. Then $(w_1, w_i) \in A^r \setminus A$ for $3 \leq i \leq r$, and the same is true with either (w_1, w_{r+1}) or (w_1, x_0) ; by property (Q), which must be satisfied with equality, there is no other arc in $A^r \setminus A$ having w_1 as an extremity ; in particular, because $r \geq 2$, the arc (x_{t-1}, w_1) must belong to A, see Figure IX. It follows that the path F which goes from x_0 to x_0 using the shortcut (x_{t-1}, w_1) is strongly connected and smaller than G^*, so we can conclude that $(x_{t-1}, x_0) \in A$, since $(x_{t-1}, x_0) \in A^r$ by (P). Let y be any vertex in

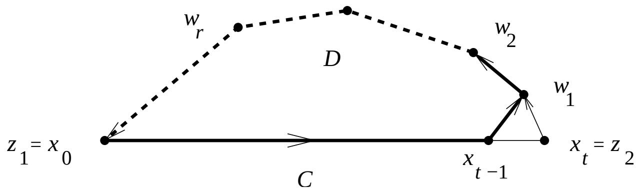


Figure IX – *The paths C and D and the path F (in bold).*

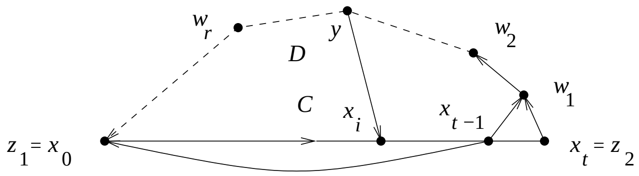


Figure X – *The paths C and D and the arc (y, x_i).*

$D \setminus \{x_0, x_t\}$. If there is an arc (y, x_i) with $0 \le i \le t-1$, see Figure X, the same argument with $x_t w_1 \ldots y x_i x_t$ shows that $(x_t, x_i) \in A$. Then we can see in Figure XI a path going from x_t to x_i, then to x_{t-1}, then to x_0, that is, a path from x_t to x_0 which uses only vertices in C; this contradicts the minimality of G^*.

So we can assume that there is no arc (y, x_i), $0 \le i \le t-1$, in A. Using again the property of the path F, this means however that all the arcs (y, x_i), $0 \le i \le t-1$, belong to $A^r \setminus A$, which represents more than $r-1$ arcs, since $t > r$. □

Thus, G^* is made of the path C, plus some arcs (x_i, x_j), $i > j$, which make G^* strongly connected. Let D be a shortest path in G^* from x_t to x_0. If $D \neq x_t x_0$, let (x_t, x_h), $h \neq 0$, be the first arc in D. Let us consider, in D, the next arc which goes "from right to left", that is, which reads (x_j, x_k) with $k < j$ and $j \ge h$, see Figure XII. Because D is a shortest path, we have $k < h$.

Since the path $x_h \ldots x_j \ldots x_t x_h$ is smaller than G^*, all its vertices are within distance r from each other, and as before we can conclude that $(x_t, x_i) \in A$ for $h \le i \le t-1$. But if $(x_t, x_i) \in A$ for $h < i \le j$, then $x_t x_i \ldots x_j x_k$ yields a path from x_t to x_0 shorter than D; so $h = j$, and (x_t, x_h) and (x_h, x_k) belong to A, which in turn implies that $(x_t, x_k) \in A$, because $r \ge 2$, again yielding a path shorter than D. Therefore we have shown that $D \neq x_t x_0$ is impossible : actually, $(x_t, x_0) \in A$.

This implies that $d_G(x_t, x_1) \le 2$, so $(x_t, x_1) \in A$, and step by step, $(x_t, x_j) \in A$ for $0 \le j \le t-1$. Similarly, $(x_i, x_j) \in A$ for $0 \le j < i \le t-1$, and we have proved the following result.

Lemma I.13. *If G achieves the bound $\Lambda(r, n)$, then G^* has vertex set $V^* = \{x_i : 0 \le i \le$*

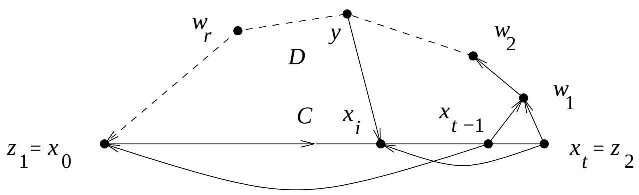


Figure XI – *A new path from x_t to x_0.*

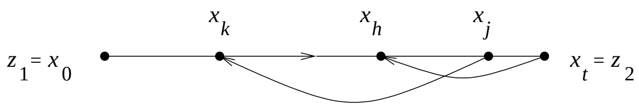


Figure XII – *Going from x_t to x_0.*

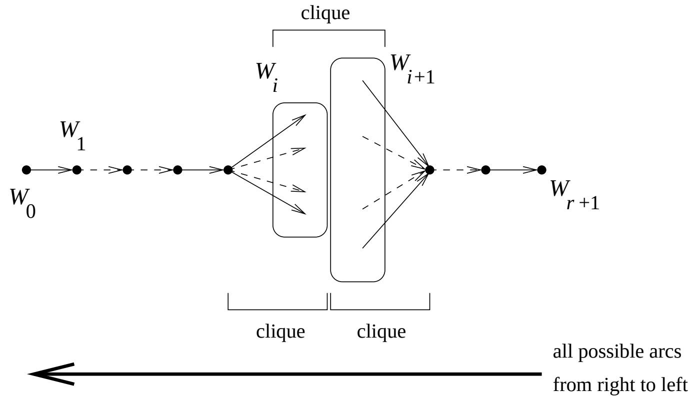


Figure XIII – *Type 1 directed graphs : $\delta = r + 1$. Any set of vertices represented by a rounded off rectangle may be reduced to a single vertex. The subscript i can vary between* 0 *and r. All arcs going from right to left exist.*

$n^* - 1\}$ and arc set given by (I.3) with $n^* = t + 1 = n_0$. □

For simplicity, we refer, for the definitions of digraphs of type 1 and type 2, to Figures XIII and XIV :

– a digraph $G = (V, A)$ is of type 1 if it has diameter $\delta = r + 1$ and if it can be represented as in Figure XIII ; note in particular that there are at most two non singleton sets, and if they exist, they are consecutive ;

– a digraph $G = (V, A)$ is of type 2, with diameter δ, if it can be represented as in Figure XIV ; if $\delta > r + 2$, there are at most four non singleton sets, two at each end ; if $\delta = r + 2$, there are at most three non singleton sets, one at one end and two at the other end ; if $\delta = r + 1$, there are at most two non singleton sets, one at each end.

Note that the graph G^* in Lemma I.13 is of type 1 with $|W_i| = |W_{i+1}| = 1$, or of type 2 with $\delta = r + 1$ and $|W_0| = |W_{r+1}| = 1$.

The digraphs of type 1 and 2 just described are very similar to the undirected graphs of type 1 and 2 used in [11, Th. 7].

Theorem I.14. *Let $G = (V, A)$ be a connected digraph of order n such that $G^r \neq K_n$ and $|A^r \setminus A| = \Lambda(r, n)$.*

Then G is of type 1 or of type 2.

Proof (abridged **A VOIR**). First, it has to be checked that these digraphs do satisfy $|A^r \setminus A| = \Lambda(r, n)$. This can be seen using the following argument : first, in all cases there are at least r singleton sets W_i ; second, if all sets W_i are singletons, i.e., we have a path, then we know that the graph meets $\Lambda(r, n)$; third, we observe that, starting from a path, if we add one by one the vertices y belonging to the sets W_j which are not singletons, each vertex y brings exactly $r - 1$ new arcs in $A^r \setminus A$, arcs (y, z) or (z, y) according to the position of the non singleton set it belongs to, with $W_i = \{z\}$ and $r \geq |i - j| \geq 2$; these arcs are counted only once ; finally, we use that $\Lambda(r, n)$ is linear in n, with the factor $r - 1$.

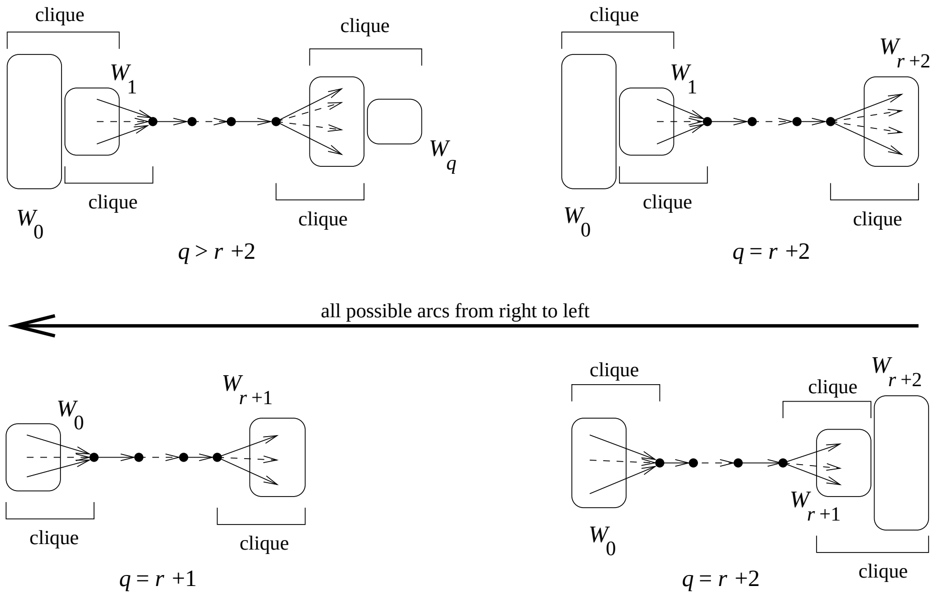


Figure XIV – *Type 2 directed graphs : $\delta > r+2$, $\delta = r+2$, or $\delta = r+1$.*

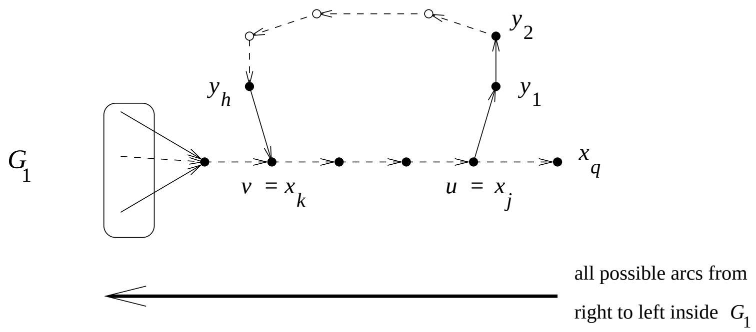


Figure XV – *The arch Y.*

Next, we consider a graph $G = (V, A)$ of order n and diameter δ which meets the bound $\Lambda(r, n)$, and an induced subgraph $G_1 = (V_1, A_1)$ of G which is of type 1 or 2 (such a graph exists, cf. Lemma I.13), and whose order is maximum. If $G_1 = G$, we are done, so we assume that $V \setminus V_1 \neq \emptyset$. As in the proof of Theorem I.4, there is a smallest arch $Y = \{y_1, y_2, \ldots, y_h\} \subseteq V \setminus V_1$ such that $H = y_1 y_2 \ldots y_h$ is a directed path in G and the arcs (u, y_1), (y_h, v) are in A, where u and v are two, non necessarily distinct, vertices in V_1 ; cf. Figure VIII. Because G_1 is of type 1 or 2, we can choose vertices $x_i \in W_i$, $0 \leq i \leq \delta$, such that $u = x_j$ and $v = x_k$ for some j, k in $\{0, 1, \ldots, \delta\}$, see Figure XV. Note that there is no arc with origin in V_1 and end y_2. We call C the path $x_0 x_1 \ldots x_\delta$, which is a shortest path from x_0 to x_δ.

We are going to show that the arch Y contains only one vertex ; suppose on the contrary that $h \geq 2$.

Lemma I.7 states that y_1 is the origin of $r-1$ arcs belonging to $A^r \setminus A$, so, if G achieves the bound, y_1 cannot be the end of any arc in $A^r \setminus A$.

Since $(x_j, y_1) \in A$, we have $(x_{j-1}, y_1\} \in A^r$ and therefore $(x_{j-1}, y_1\} \in A$; step by step, we obtain that for all i between 0 and $j-1$, $(x_i, y_1) \in A$. For $i > j$, the arc (x_i, x_j) is in A, so as before we must have $(x_i, y_1) \in A$: we have just proved that all arcs (x_i, y_1),

$0 \le i \le \delta$, are in A, which implies that $d_G(x_i, y_2) \le 2$. So $(x_i, y_2) \in A^r$, and we have observed that $(x_i, y_2) \notin A$. This represents $\delta + 1 > r - 1$ arcs in $A^r \setminus A$ with origin in V_1 and end y_2, which is more than stated in Lemma I.8, and so the bound cannot be achieved. So the assumption $h \ge 2$ led to a contradiction, and we have, setting $y = y_1$, $Y = \{y\}$. By Lemma I.10 and to satisfy the bound with equality, we have in $A^r \setminus A$ exactly $r - 1$ arcs with end or origin y.

Let h be the smallest subscript such that $(x_h, y) \in A$ and k be the largest subscript such that $(y, x_k) \in A$; then in $A^r \setminus A$, there are the $\min\{r-1, h\}$ arcs (x_{h-1}, y), (x_{h-2}, y), $\ldots$ and the $\min\{r-1, \delta-k\}$ arcs $(y, x_{k+1}), (y, x_{k+2}), \ldots$

Since C is a shortest path, we cannot have shortcuts going through y, which means that the existence of (x_h, y) in A implies the nonexistence in A of $(y, x_{h+3}), (y, x_{h+4}), \ldots$, and we have $k \le h + 2$.

Therefore, in $A^r \setminus A$, we have at least

$$\Gamma = \min\{r-1, h\} + \min\{r-1, \delta - h - 2\}$$

arcs with one extremity in C and one extremity on y, and so the integer Γ must be less than or equal to $r - 1$.

(i) If $h \le r - 1$ and $\delta - h - 2 \le r - 1$, then $\Gamma = \delta - 2$, which implies $\delta = r + 1$.

(ii) If $h > r - 1$ and $\delta - h - 2 \le r - 1$, then $\Gamma = (r-1) + \delta - h - 2$, which implies $h \ge \delta - 2$.

(iii) If $h \le r - 1$ and $\delta - h - 2 > r - 1$, then $\Gamma = h + (r - 1)$, which implies $h = 0$.

The fourth case is impossible since we would have $\Gamma = 2(r-1) > r - 1$ for $r \ge 2$.

Starting from the graph G_1 which is of type 1 or 2 and has maximum order n_1, we are led to add one vertex y, with different conditions on h, r and δ, according to the three cases (i)–(iii), in order to retrieve the graph G which meets the bound $\Lambda(r, n)$. It is then tedious but straightforward to see that the resulting graph is necessarily still of type 1 or 2, contradicting the maximality of n_1.

If we are in case (i), then $\delta = r + 1$ and G_1 can be of type 1 or 2. ... ... **A** ($\pm$) **FINIR** □

I.5 Conclusion

We have addressed three problems on the sizes of G and G^r.

In Section I.2, we have given the maximum size of a strongly connected digraph with given order and diameter, and characterized the graphs meeting the bound.

In Section I.3, we have determined the minimum size of the r-th power of a strongly connected digraph, and proved that the circuit is the only graph achieving this bound.

In Section I.4, we have answered the following question : given a strongly connected digraph G of order n such that $G^r \neq K_n$, what is the minimum number of arcs that are added when going from G to G^r, and which are the graphs achieving this bound ?

Annexe J

On the Square Roots of a Graph

David Auger [1], **Irène Charon** [1],
Olivier Hudry [1], **Antoine Lobstein** [2]
{*david.auger, irene.charon, olivier.hudry, antoine.lobstein*}*@telecom-paristech.fr*

Abstract

The square of an undirected graph G is the graph G^2 on the same vertex set as G, where two vertices are adjacent if and only if their distance in G is 1 or 2. It is known that deciding whether a graph is a square is NP-complete, but this problem can be solved in polynomial time if one only searches for square roots in the class of trees, or in the class of bipartite graphs. In a first part, we extend these results by giving a polynomial algorithm which decides if a graph admits square roots in the class $\mathcal{G}_6$ of connected graphs with girth at least 6, and outputs them in case of a positive answer. In a second part we study the maximal number of square roots of a graph. We show that a connected graph on n vertices with radius at least 2 admits at most one square root with girth at least 7, whereas its number of square roots in $\mathcal{G}_6$ is bounded by $\frac{1}{2} + \sqrt{\frac{1}{4} + 2n}$, all these roots being isomorphic. We show that this bound is tight for infinitely many values of n and we give the graphs which attain it.

Keywords : Powers of Graphs, Roots of Graphs, Distances in Graphs.

J.1 Motivation and notation

By *graph* we mean a non-empty, finite and undirected graph, without loops nor multiple edges. If G is a graph, we denote respectively by $V(G)$ and $E(G)$ the sets of vertices and edges of G. If x, y are vertices of G, we simply denote an edge by xy. If $x \in V(G)$, a *neighbour* of x is a vertex $y \in V(G)$ such that $xy \in E(G)$. The *neighbourhood* of x in G is denoted by $N_G(x)$; it is the set of all neighbours of x. For further basic terminology about graphs used in this paper such as paths, cliques, cycles, stables sets, or connectivity, we

2. Institut TELECOM - TELECOM ParisTech & Centre National de la Recherche Scientifique - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13 - France

2. Centre National de la Recherche Scientifique - LTCI UMR 5141 & Institut TELECOM - TELECOM ParisTech, 46, rue Barrault, 75634 Paris Cedex 13 - France

refer to [21], whereas for the notion of NP-completeness and general background about algorithmic complexity we refer to [50].

The *square* of an undirected graph G is the graph G^2 whose vertex set is $V(G)$ and such that for all distinct vertices x and y, the edge xy belongs to $E(G^2)$ if and only if $d_G(x, y) \leq 2$, where $d_G(x, y)$ denotes the usual distance between x and y in G, i.e. the length of a shortest path in G between x and y (or $+\infty$ if there is none). Obviously, G is a subgraph of G^2, and we obtain $E(G^2)$ by adding to $E(G)$ all the edges $xy \notin E(G)$ such that there is a vertex z with $xz \in E(G)$ and $zy \in E(G)$.

A graph H is a *square root* of a graph G if $H^2 = G$. Note that in this case H is connected if and only if G is connected.

It is convenient to consider the square G^2 of a graph G (and more generally *powers* G^p of G) when dealing with problems related to the distance 2 (respectively p) in G. For instance, looking for a set of vertices whose pairwise distances are at least 3 in G amounts to finding a stable set in G^2. Our own interest came from similar observations : a 2-distance-dominating set in G is a dominating set in G^2 (see [57]), and a 2-identifying code in G is an identifying code in G^2 (see for instance [38] or more generally [73]). In addition to this, squares of graphs have hamiltonian properties : for instance, Fleishner proved in 1974 that the square of a 2-vertex-connected graph admits an Hamilton cycle (see [47]). It follows that the recognition of squares of graphs, and the computation of their square roots, are natural and interesting problems. Several results are already known : Mukhopadhyay gave a characterization of graphs admitting a square root ([78]), Ross and Harary proved that up to isomorphism, a graph can have at most one square root in the class of trees ([85]), whereas Lin and Skiena gave a linear algorithm for recognizing squares of trees ([72]) and computing their tree square root. On the other hand, Motwani and Sundan proved the NP-completeness of the recognition of squares of graphs in the general case, and conjectured that recognizing squares of bipartite graphs was also NP-complete ([77]) but Lau showed that this is not the case in [71].

In this paper, we study graphs admitting square roots with no cycles of small length. Recall that the *girth* of a graph is the length of a shortest cycle in G ; if there is no cycle in G the girth is considered to be infinite. For $k \geq 3$ (respectively $k = \infty$), we denote by $\mathcal{G}_k$ the class of all connected graphs with girth at least k (respectively the class of trees). Let us explain briefly how the paper is organized :

- in section J.2, along with a few lemmas about graphs admitting square roots in $\mathcal{G}_6$, we exhibit a polynomial algorithm for the computation of these roots ; this ends the algorithmic part of the paper ;
- in section J.3, we prove that apart from degenerate cases a graph admits at most one square root (as a subgraph, i.e. *not* up to isomorphism) in $\mathcal{G}_7$;
- in sections J.4, J.5 and J.6 we focus on the maximal number of roots in $\mathcal{G}_6$. In section J.4, we prove that if a graph may admit several distinct square roots in $\mathcal{G}_6$, all these roots must be isomorphic, and we give a canonical isomorphism between these roots. In section J.5 we construct two families of graphs $\{R_\alpha\}_{\alpha \geq 4}$ and $\{R_{\alpha,\beta}\}_{\alpha,\beta \geq 2}$, having multiple square roots in $\mathcal{G}_6$. In the last section, we show that a graph admitting multiple square roots in $\mathcal{G}_6$ can be decomposed in blocks isomorphic to graphs of the families $\{R_\alpha\}_{\alpha \geq 4}$ and $\{R_{\alpha,\beta}\}_{\alpha,\beta \geq 2}$. This enables us to prove that (apart from degenerate cases) the maximal number of square roots in $\mathcal{G}_6$ of a graph G is bounded above by
$$\frac{1}{2} + \sqrt{\frac{1}{4} + 2\,|V(G)|},$$
and that this bound it tight for infinitely values of n.

Let us precise a little more notation before beginning. We will distinguish *equal* graphs (in the sense of equal labelled graphs) from *isomorphic* ones. We consider two graphs G_1 and G_2 to be equal if and only if $V(G_1) = V(G_2)$ and $E(G_1) = E(G_2)$ (let us insist that the sets of vertices must be identical) ; whereas G_1 and G_2 will be *isomorphic* if one can find a bijective function f from $V(G_1)$ to $V(G_2)$, such that $f(x)f(y) \in E(G_2)$ if and only if $xy \in E(G_1)$.

If G is a graph, for all $x \in V(G)$ and $k \geq 0$ we denote by $B_G(x,k)$ (respectively by $\Gamma_G(x,k)$) the *ball* (resp. the *sphere*) of centre x and radius k, i.e. the set of all vertices $y \in V(G)$ such that $d_G(x,y) \leq k$ (respectively $d_G(x,y) = k$).

If V is a set (of vertices) and $x \in V$, the *star* with centre x is the graph with vertex set V whose edges are all the xy for every $y \in V - x$; here $V - x$ denotes $V \setminus \{x\}$.

It will be convenient to write $A + B$ for the union $A \cup B$ of two sets A and B, or $A - B$ for $A \setminus B$. For instance, $A + a - (b + C)$ is a short notation for $(A \cup \{a\}) \setminus (\{b\} \cup C)$.

Finally, if $n \geq 1$ we denote by $[n]$ the set $\{1, 2, \cdots n\}$, and if $0 \leq k \leq n$ the the set of k-element subsets of $[n]$ will be denoted by $\binom{[n]}{k}$.

J.2 Computation of square roots in $\mathcal{G}_6$

Let now $k \geq 3$ or $k = \infty$ and consider the following problem :

Problem Π_k : Given a connected graph G, is there a graph $H \in \mathcal{G}_k$ such that $H^2 = G$?

Problem Π_k belongs to the class P of problems that can be solved in polynomial time for $k = \infty$ by the result of Lin and Skienna ([72]), whereas it is NP-complete for $k = 3$ (see [77]). Let us state the main result of this section :

Theorem J.1. *There is a polynomial algorithm which determines if a graph G admits square roots in $\mathcal{G}_6$ and computes them in case of positive answer.*

Since it is polynomial to compute the girth of any graph, Theorem J.1 shows that Π_k belongs to P for all $k \geq 6$, and in particular for $k = \infty$. Let us also note that we can without loss of generality restrict ourselves to connected graphs.

The principle of the algorithm is to use necessary conditions for a graph $H \in \mathcal{G}_6$ to be a square root of G, in order to compute a list of graphs $H_1, H_2, \cdots, H_l$ containing all possible square roots $H \in \mathcal{G}_6$ of G. When this is done, we just have to compute H_i^2 for each $1 \leq i \leq l$ and compare it to G : thus we obtain all the square roots of G which belong to $\mathcal{G}_6$. To do this, first we will consider the case where G is a complete graph ; in this case it is easy to list the square roots of G (Lemma J.4). If G is not complete, we will consider a vertex x of G, and show that every square root $H \in \mathcal{G}_6$ of G can be computed in polynomial time if we know the neighbourhood $N_H(x)$ of x in H (Lemma J.5). Thus if we show that we can compute in polynomial time every possible neighbourhood of x in the square roots, we will be done. We will give (rough) evaluations of the complexities of our algorithms to ensure its polynomiality, and for this we suppose that G is coded by its adjacency matrix.

Let us begin with two simple but useful lemmas :

Lemma J.2. *Let $H \in \mathcal{G}_6$ and $x_0x_1x_2x_3$ be a path of length 3 in H. Then $d_H(x_0, x_3) = 3$.*

Proof. Clearly $d_H(x_0, x_3) \leq 3$, and we cannot have $d_H(x_0, x_3) = 1$ because then $x_0x_1x_2x_3x_0$ would be a cycle of length 4 in H, which would contradict $H \in \mathcal{G}_6$. Suppose now that $d_H(x_0, x_3) = 2$; then there must exist a vertex y such that x_0y and yx_3 belong to $E(H)$. On the one hand, we cannot have $y = x_1$ or $y = x_2$, which would imply the existence of a triangle in H ($x_1x_2x_3x_1$ or $x_0x_1x_2x_0$) ; but on the other hand if y is distinct from x_1 and x_2 then $x_0x_1x_2x_3yx_0$ is a cycle of length 5 : a contradiction. Thus $d_H(x_0, x_3) = 3$. □

Lemma J.3. *(Adjacency Lemma) Let $H \in \mathcal{G}_6$ and $xy \in E(H)$. Let $z \in V(H)$, distinct from x and y, such that $d_H(z, x) \leq 2$ and $d_H(z, y) \leq 2$. Then $zx \in E(H)$ or $zy \in E(H)$.*

Proof. Suppose that neither xz nor yz belong to $E(H)$. Then we have $d_H(x, z) = 2$ and $d_H(y, z) = 2$, and so there are vertices x' and y' such that the edges xx', $x'z$, yy' and $y'z$ all belong to $E(H)$. If $x' = y'$, then $xyx'x$ is a cycle in H of length 3, whereas if $x' \neq y'$ we have a cycle $xyy'zx'x$ of length 5. In both cases, we cannot have $H \in \mathcal{G}_6$. □

We now examine the case of complete graphs. For $n \geq 1$, the complete graph K_n is the graph with vertex set $[n]$ and whose edges are all the pairs xy with $x \neq y \in [n]$. More generally, a *complete* graph is a graph isomorphic to K_n for some $n \geq 1$.

Lemma J.4. *Let $n \geq 3$. The complete graph K_n admits exactly n square roots in $\mathcal{G}_6$, which are the stars centred on each one of the n vertices of K_n.*

Proof. Clearly the n stars are square roots of K_n, and belong to $\mathcal{G}_6$. Suppose now that $H \in \mathcal{G}_6$ is a square root of K_n. Since H^2 is complete, the distance between two vertices is at most 2 ; so by Lemma J.2 the graph H contains no path of length 3.

Now H cannot be complete since it contains no triangles, but it is connected, so there must exist two vertices x, z such that $d_H(x, z) = 2$. Let $y \in V(H)$ such that xyz is a path from x to z. Since there are no paths of length 3 in H, the vertices x and y must have degree 1 in H and no vertex can be at distance 2 from y. We conclude that every vertex $v \neq y$ must satisfy $d_H(v, y) = 1$. Since H contains no triangles, it is a star with centre y. □

The next lemma shows that given a graph G and a vertex x, there is at most one square root $H \in \mathcal{G}_6$ of G where x has a specified neighbourhood, and that this square root can be computed in polynomial time.

Lemma J.5. *Let G be a connected graph, $x \in V(G)$ and $A \subset V(G) - x$. There is a polynomial algorithm which, given G, x and A, computes a graph $f(G, x, A)$ which belongs to $\mathcal{G}_6$, such that if G admits a square root $H \in \mathcal{G}_6$ with $N_H(x) = A$, then $H = f(G, x, A)$. In particular, G admits at most one square root H in $\mathcal{G}_6$ such that $N_H(x) = A$.*

Proof. Let G, x and A be given. Suppose that G admits a square root $H \in \mathcal{G}_6$ with $N_H(x) = A$, and let $y \in A$. It suffices to show that $N_H(y)$ can be computed in polynomial time ; then one can also compute the neighbourhoods of neighbours of y, and so on : since G is connected we can compute the neighbourhood in H of every vertex by a simple search algorithm.

First note that y cannot have a neighour z in A, because $xyzx$ would be a cycle of length 3 in H. Thus all the neighbours $z \neq x$ of y satisfy $d_H(x, z) = 2$; we also have $d_H(y, z) = 1 \leq 2$. Conversely, if $d_H(x, z) = 2$ and $d_H(y, z) \leq 2$, consider a path xtz from x to z ; if $t \neq y$, then $yxtz$ is a path in H so by Lemma J.2 we must have $d_H(y, z) = 3$, a contradiction. So $t = y$ and thus $z \in N_H(y)$. We have shown that

$$\begin{aligned} N_H(y) &= x + \Gamma_H(x,2) \cap B_H(y,2) \\ &= x + (N_G(x) - A) \cap N_G(y) \end{aligned}$$

and so the neighbourhood of y in H can be computed if we know G, x and the neighbourhood A of x in H. □

Note that if G has n vertices, the previous algorithm runs in time $O(n^2)$: there are n computation of neighbourhoods, and each one needs $O(n)$ operations ; the search algorithm also requires $O(n^2)$ operations.

Now that we can build roots from neighbourhoods, we need to find all the possible neighbourhoods of a chosen vertex x. First, we consider the case where x could have a single neighbour in H :

Lemma J.6. *Let $H \in \mathcal{G}_6$ with $H^2 = G$, and suppose that G is not a complete graph. Then a vertex x has degree one in H if and only if $N_G(x)$ is a clique of G.*

Proof. The "only if" part being trivial, let us suppose that x admits (at least) two neighbours y and z. Since H cannot be a star, one neighbour of x has another neighbour $t \neq x$ in H. So without loss of generality, we can suppose that $zxyt$ is a path in H. By Lemma J.2, we have $d_H(z,t) = 3$, thus $N_G(x)$ is not a clique in G. □

If we find no vertex x in G such that $N_G(x)$ is a clique, then every vertex must have a degree at least two in every square root $H \in \mathcal{G}_6$ of G. For a given x this conducts, a priori, to a number of possible neighbourhoods which could be exponential in the size of G (all the possible subsets of $N_G(x)$) ; however, the next lemma will show that this is never the case.

Lemma J.7. *Let $H \in \mathcal{G}_6$, $x \in V(H)$ and y, z be two distinct neighbours of x. Then*

$$B_H(x,1) = B_H(x,2) \cap B_H(y,2) \cap B_H(z,2).$$

Proof. Let $B = B_H(x,2) \cap B_H(y,2) \cap B_H(z,2)$. Clearly, $B_H(x,1) \subset B$. For the other inclusion, consider $u \in B$; if $u = x, y$ or z then $u \in B_H(x,1)$, so suppose that $u \notin \{x,y,z\}$. Then as $xy \in E(H)$, $d_H(x,u) \leq 2$ and $d_H(y,u) \leq 2$ we must have xu or yu in $E(H)$ by the adjacency Lemma J.3 ; and for symmetric reasons we have xu or zu in H. Thus xu belongs to $E(H)$, otherwise we would have $yu \in E(H)$ and $zu \in E(H)$, and in this case $uyxzu$ would be a cycle of length 4 in H. We conclude that $u \in B_H(x,1)$. □

In other words, the previous lemma shows that

$$B_H(x,1) = B_G(x,1) \cap B_G(y,1) \cap B_G(z,1),$$

and so specifying two neighbours of x in H is sufficient to compute its whole neighbourhood.

Now we will use the previous lemmas in order to describe the algorithm mentioned in Theorem J.1. Since our interest is theoretical, we prefer to explain how the algorithm works rather than giving its pseudo-code.

Proof of Theorem J.1. Let G be given by its adjacency matrix ; we suppose that G is connected, has n vertices and minimal degree δ_{min}. The algorithm runs in five phases ; at the end of each phase the algorithm can stop.

Phase 1 : Preliminaries

Find the minimal degree δ_{min} of G and a vertex x with minimal degree. This is done in time $O(n^2)$. Go to Phase 2.

Phase 2 : Does $n \leq 2$?

If this is the case G is its only square root : exit the algorithm, otherwise go to the Phase 3.

Phase 3 : Is G a complete graph ?

This is the case if $\delta_{min} = n - 1$; in case of positive answer, we know by Lemma J.4 that G admits as square roots in $\mathcal{G}_6$ the n stars on $V(G)$ and the algorithm terminates. Otherwise, go to Phase 4.

Phase 4 : Are there square roots where x has degree one ?

We know by Lemma J.6 that this case can happen only if $N_G(x)$ is a clique, so we begin by this test ; if $N_G(x)$ is not a clique, we go directly to Phase 5, otherwise we continue and the algorithm stops at the end of Phase 4.

Suppose that $N_G(x)$ is a clique. Then for every choice of $y \in N_G(x)$, we can compute by the algorithm of Lemma J.5 a graph H which is the only candidate to be a square root of G where the only neighbour of x is y ; then we have to check if $H^2 = G$ or not. Thus we obtain all the square roots of G in $\mathcal{G}_6$ and we stop the algorithm. Since the algorithm of Lemma J.5 runs in time $O(n^2)$ and we can check if $H^2 = G$ in time $O(n^2)$, and since there are δ_{min} choices of possible neighbours for x, this phase runs in the worst case in time $O(\delta_{min} n^2)$.

Phase 5 : Are there square roots where x has degree at least two ?

If we get to this phase we know that $n \geq 3$, that G is not complete and furthermore $N_G(x)$ is not a clique. If $H \in \mathcal{G}_6$ is a square root of G, then x must have degree at least two in H ; so if we choose two possible neighbours y and z in a square root H (there are $O(\delta_{min}^2)$ choices) we can compute the neighbourhood $N_H(x)$ by Lemma J.7 and then H by Lemma J.5 ; then we can check if $H^2 = G$ or not : thus we get the list of all square roots of G in time $O(\delta_{min}^2 n^2)$. □

J.3 Uniqueness of the square roots in $\mathcal{G}_7$

Here we extend to $\mathcal{G}_7$ the result of Ross and Harary ([85]) previously mentioned which states that a graph admits at most one square root in the class of trees ; this result is *up to isomorphism.* Since we showed in Lemma J.4 that a complete graph on n vertices admits n square roots in $\mathcal{G}_6$, which in fact are trees, we cannot claim that square roots in $\mathcal{G}_7$, considered as subgraphs rather than up to isomorphism, are always unique : we need to exclude the degenerate case of graphs with radius one (or zero). Recall that the radius of a graph is the least $r \geq 0$ such that there is $v \in V(G)$ such that $B_G(v, r) = V(G)$. We will restrict ourselves to graphs G with radius at least 2, i.e. graphs where every vertex admits a vertex at distance 2. First we consider the case of graphs with radius at most 1 :

Lemma J.8. *Let G be a connected graph with radius at most* 1*, admitting r square roots in $\mathcal{G}_6$, where $r \geq 2$. Then either :*

– *G is a complete graph and $r = |V(G)|$;*

– *or G is the union of two complete graphs with a common edge, i.e. $V(G)$ can be decomposed in $V(G) = V_1 \cup V_2$, where $|V_1| \geq 3$ and $|V_2| \geq 3$ and $|V_1 \cap V_2| = 2$, and all the edges of G have both ends in V_1, or both ends in V_2 (see Fig.I). In this case, $r = 2$.*

In both case, all square roots of G are isomorphic.

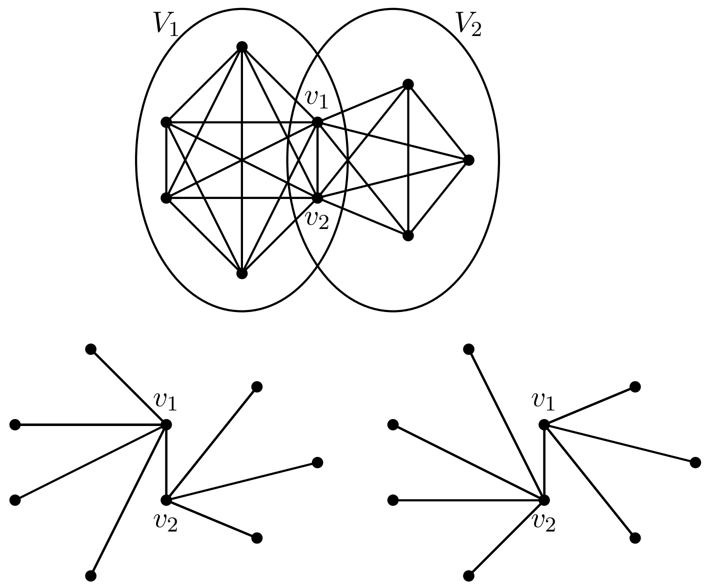


Figure I – *A graph G which is the union of two complete graphs, and its two square roots in $\mathcal{G}_7$*

Proof. If G is a complete graph, then the result follows from Lemma J.4. If it is not a complete graph, then the result easily follows from Lemma J.12 which will be proved in the next section. We leave this to the reader. □

Now we exclude the case of graphs with radius 1 or 0 :

Theorem J.9. *A connected graph with radius at least 2 admits at most one square root in $\mathcal{G}_7$.*

Let us begin with a lemma :

Lemma J.10. *Let $H \in \mathcal{G}_7$. Let x, y be two distinct vertices with degree at least 2 in H, and suppose that $d_H(x, y) \leq 2$. Then $d_H(x, y) = 2$ if and only if $B_H(x, 2) \cap B_H(y, 2)$ is a clique in H^2.*

Proof. Consider x, y as in the statement of the lemma. If $d_H(x, y) = 2$, then there is a single path xzy in H and we claim that $B_H(x, 2) \cap B_H(y, 2) = B_H(z, 1)$, which is a clique in H^2. Indeed z and its neighbours clearly belong to $B_H(x, 2) \cap B_H(y, 2)$; conversely, suppose that $t \in B_H(x, 2) \cap B_H(y, 2)$ and $t \notin B_H(z, 1)$. Then a path from x to t, followed by a path from t to y, followed by yzx, yields a cycle of length at most 6, which contradicts $H \in \mathcal{G}_7$.

Suppose now that $d_H(x, y) = 1$. Then since the degrees of x and y are at least two, there must exist a path $x'xyy'$ in H where x' and y' are respective neighbours of x and y. We have $x', y' \in B_H(x, 2) \cap B_H(y, 2)$, but $d_H(x', y') = 3$ by Lemma J.2, and so $B_H(x, 2) \cap B_H(y, 2)$ is not a clique in H^2. □

Proof of Theorem J.9. Let a connected graph G with radius at least 2 be given, and suppose that $G = H^2$ where $H \in \mathcal{G}_7$. We consider two types of vertices : the set $\mathcal{L}$ of vertices which have degree one in H (by Lemma J.6, $\mathcal{L}$ is the set of vertices x such that $N_G(x)$ is a clique in G) and the remaining set $V(G) - \mathcal{L}$ of vertices which have degree at least 2 in H. Let us characterize the edges in $E(G)$ which also belong to $E(H)$, and thus show the unicity of H.

First, consider an edge $xy \in E(G)$ with x and y in $V(G) - \mathcal{L}$. By Lemma J.10 we have $xy \in E(H)$ if and only if $B_H(x, 2) \cap B_H(y, 2)$ is a clique in H^2.

Consider now an edge $xz \in E(G)$ with $x \in \mathcal{L}$. Since $x \in \mathcal{L}$, it admits exactly one neighbour in H ; we show that there is only one possibility for this neighbour and thus settle the case of the edge xz. So let

$$N_G(x) = \{z_0, z_1, \cdots, z_k\}$$

be the neighbourhood of x in G ; then the neighbour of x in H must be one of the z_i's. We claim that if we compute the set

$$\mathcal{H} = \{N_G(y) \cap N_G(x) : y \notin N_G(x)\}$$

then there must exist in $\mathcal{H}$ two sets of the form

$$\{z_i\} \subsetneq \{z_i, z_j\} \tag{J.1}$$

and then z_j must be the neighbour of x in H.

To see this, suppose without loss of generality that the neighbour of x in H is z_0 ; then $z_1, z_2, \cdots, z_k$ must be the other neighbours of z_0 (see Fig. II).

It is easily seen that since $H \in \mathcal{G}_7$, if $A \in \mathcal{H}$ is nonempty we must have $A = \{z_i\}$ or $A = \{z_0, z_i\}$ for a $1 \leq i \leq k$ and so (J.1) characterizes $z_0 = z_j$. Let us show that sets verifying (J.1) in $\mathcal{H}$ can be found : since G has radius at least two there must exist a vertex v such that $d_H(z_0, v) = 3$, so we can suppose that there is a path $z_0 z_1 u v$ in H (see Fig. II). Then we have

$$N_G(u) \cap N_G(x) = \{z_0, z_1\}$$

and

$$N_G(v) \cap N_G(x) = \{z_1\}.$$

and these sets belong to $\mathcal{H}$; thus z_0 can be identified. □

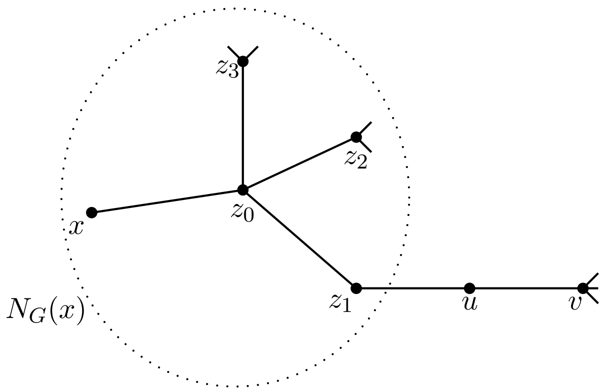


Figure II – *Proof of Theorem J.9*

J.4 The structure of coroots

We say that two distinct graphs H_1 and H_2 are *coroots* if $V(H_1) = V(H_2)$ and $H_1^2 = H_2^2$; in other words H_1 and H_2 are distinct square roots of the same graph. We have seen that a graph with radius at least 2 admits at most one square root in $\mathcal{G}_7$; however, such a graph may have several square roots in $\mathcal{G}_6$. In this case the different square roots are closely related and the whole graph must have a particular structure. In this part we will show :

Theorem J.11. *If a graph admits several square roots in $\mathcal{G}_6$ they are all isomorphic. More precisely, if $H_1, H_2 \in \mathcal{G}_6$ are distinct coroots and if D denotes $E(H_1) \cap E(H_2)$, the map $f : V \longrightarrow V$, such that $f(x) = y$ and $f(y) = x$ for all vertices x, y with $xy \in D$, and $f(z) = z$ for other vertices, is well defined and is an isomorphism from H_1 to H_2.*

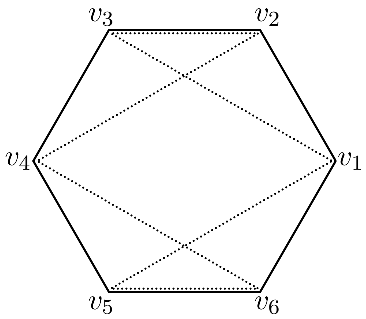


Figure III – *a cycle $v_1v_2v_3v_4v_5v_6v_1$, and one of its three coroots, $v_1v_3v_2v_4v_6v_5v_1$.*

Let us first consider the simple exemple of the cycle C_6, i.e. the graph with vertices v_i for $i \in \{1, 2, \cdots, 6\}$ and edge set

$$E(C_6) = \{v_iv_{i+1}, i = 1, 2, \cdots, 5\} \cup \{v_6v_1\}.$$

This graph admits exactly three coroots in $\mathcal{G}_6$, which can be obtained by switching vertices in opposite edges of the cycle (see fig. III). In other words, the square C_6^2 of the cycle C_6 admits four square roots in $\mathcal{G}_6$.

Let G be a graph admitting two distinct square roots H_1 and H_2. We define a *double edge* of the pair $\{H_1, H_2\}$ to be an edge of G belonging to both H_1 and H_2, i.e. the set of double edges of this pair is

$$D = E(H_1) \cap E(H_2).$$

We consider for the next lemma the following point of view : if we know that $H_1 \in \mathcal{G}_6$ admits a coroot $H_2 \in \mathcal{G}_6$, then the set $D = E(H_1) \cap E(H_2)$ of double edges can be considered as a subset of $E(H_1)$: we will show that D is a maximal matching in H_1, with additional properties. Conversely, if we find a set $D \subset E(H_1)$ satisfying these properties, then it must be the set of double edges for H_1 and a coroot $H_2 \in \mathcal{G}_6$. Now we state this more precisely :

Lemma J.12. *Let $H_1 \in \mathcal{G}_6$, and $D \subset E(H_1)$. Then there is a coroot $H_2 \in \mathcal{G}_6$ of H_1 such that $D = E(H_1) \cap E(H_2)$ if and only if D satisfies the three following conditions :*
(i) there are no adjacent edges in D ;
(ii) if $e \in E(H_1)$ but $e \notin D$, then e is adjacent to exactly one edge $e' \in D$;
(iii) if $abca'b'$ is a path in H_1 with ab and $a'b'$ in D, then there is a vertex $c' \in V(H_1)$, distinct from a, b, c, a' and b', such that $b'c'$ and $c'a$ belong to $E(H_1)$, so that $abca'b'c'a$ is a cycle of length 6 in H_1 (see fig. IV).

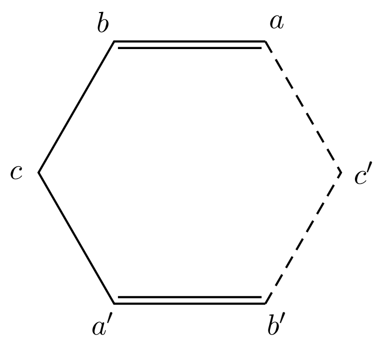


Figure IV − *condition (iii) in Lemma J.12*

Proof. Let H_1 and H_2 be coroots in $\mathcal{G}_6$ and let $D = E(H_1) \cap E(H_2)$.

Proof of (i). Suppose that x, y, z are three distinct vertices such that $xy \in D$ and $yz \in D$. We prove that $N_{H_1}(y) = N_{H_2}(y)$ and Lemma J.5 will imply that $H_1 = H_2$. Let a be a neighbour of y in H_1. We have $d_{H_1}(a, x) \leq 2$ and $d_{H_1}(a, y) \leq 2$, and so since $H_1^2 = H_2^2$ we must also have $d_{H_2}(a, x) \leq 2$ and $d_{H_2}(a, y) \leq 2$. So by the Adjacency Lemma J.3, since the edge xy belongs to $E(H_2)$, the graph H_2 must contain one of the two edges ax or ay. By the same argument, H_2 contains one of the two edges az or ay. Suppose that $ay \notin E(H_2)$; then $ax \in E(H_2)$ and $az \in E(H_2)$. So $xyzax$ is a cycle of length 4 in H_2 : a contradiction. We conclude that $ay \in E(H_2)$, and so a is also a neighbour of y in H_2. The same conclusion holds for every neighbour of y in H_1 and thus $N_{H_1}(y) \subset N_{H_2}(y)$. For the converse inclusion, we reverse the argument.

Proof of (ii). Consider now $e = xy \in E(H_1) \setminus D$; then as H_1 and H_2 are coroots and $d_{H_1}(x, y) = 1$, we must have $d_{H_2}(x, y) = 2$, so there must exist a vertex z such that xzy is a path in H_2. From this, we deduce that $d_{H_1}(z, x) \leq 2$ and $d_{H_1}(z, y) \leq 2$. By the Adjacency Lemma J.3, applied in H_1, we have $zx \in E(H_1)$ or $zy \in E(H_1)$ and so either zx or zy belongs to $E(H_1) \cap E(H_2)$. Therefore there is an edge $e' \in D$ adjacent to e. Let us show now the uniqueness of this edge e'. By the previous result (i), as H_1 and H_2 are distinct, there cannot be two edges of D incident to x, or in y. Suppose that there are two edges $x'x$ and $y'y$ of D, adjacent to e respectively in x and in y. Then we have $x'x \in E(H_2)$, $d_{H_2}(y, x') \leq 2$ and $d_{H_2}(y, x) \leq 2$ so by Lemma J.3 we deduce that yx or yx' belongs to $E(H_2)$. We supposed $yx \notin D$, and so $yx' \in E(H_2)$. By similar arguments we also have $xy' \in E(H_2)$, and this implies that $xx'yy'x$ is a cycle of length 4 in H_2, which is impossible.

Proof of (iii). Finally, consider $abca'b'$ as stated in (iii). By the Adjacency Lemma J.3, as abc is a path in H_1 with $ab \in E(H_2)$, we must have $ac \in E(H_2)$, and by the same argument $b'c \in E(H_2)$. Therefore acb' is a path in H_2 and so in H_1 the vertices a and b' must be at distance 1 or 2 ; but they cannot be at distance 1, because $abca'b'a$ would be a cycle of length 5 in H_1. So there must exist $c' \in V(H_1)$, distinct from a and b', such that $b'c'$ and $c'a$ belong to $E(H_1)$. It is also easy to check that c' must be different from b, c and a' because it would imply the existence of short cycles.

Now we prove that if $D \subset E(H_1)$ satisfies (i), (ii) and (iii), then we can find a coroot H_2 of H_1 such that $D = E(H_1) \cap E(H_2)$.

By assumption, each vertex $v \in E(H_1)$ is adjacent to at most one edge from D. Let us define a map f from the vertex set $V(H_1)$ to itself in the following way : for each $x \in V(H_1)$, if there is $y \in V$ such that $xy \in D$, then we set $f(x) = y$; if not, we set $f(x) = x$. By (i) and (ii), it is easy to see that f is well defined and is an involution map of $V(H_1)$, i.e. f is bijective and $f^{-1} = f$.

Let us now define a graph H_2 on the same vertex set $V(H_2) := V(H_1)$, by letting $xy \in E(H_2)$ if and only if $f(x)f(y) \in E(H_1)$. By its definition H_2 is isomorphic to H_1 ; furthermore we will show that $H_1^2 = H_2^2$. This will also show Theorem J.11.

To see this, consider two distinct vertices x and y. First, we show that if $d_{H_1}(x, y) \leq 2$ then $d_{H_2}(x, y) \leq 2$.

- if $d_{H_1}(x, y) = 1$ and $xy \in D$, then $f(x) = y$ and $f(y) = x$ so $xy \in E(H_2)$ and $d_{H_2}(x, y) = 1$;
- if $d_{H_1}(x, y) = 1$ and $xy \notin D$, then x or y is adjacent to an edge of D : suppose without loss of generality that $yz \in D$ with $z \in V(H_1)$. Then $f(x) = x$, $f(y) = z$ and $f(z) = y$ and so xz and zy belong to $E(H_2)$: therefore $d_{H_2}(x, y) = 2$;
- if $d_{H_1}(x, y) = 2$, there is a path xzy in H_1. If $xz \in D$, then we have $xy \in E(H_2)$ and so $d_{H_2}(x, y) = 1$; the case $zy \in D$ is the same ;
- if $d_{H_1}(x, y) = 2$ and xyz is a path in H_1, with neither xz nor zy in D, then there must be an edge of D adjacent to xz in x or in z. In the latter case, there is a vertex t with $tz \in D$, therefore tx and ty belong to $E(H_2)$ and so $d_{H_2}(x, y) = 2$. In the former case, there must be an edge $xy' \in D$ and also an edge $yx' \in D$ by (i) and (ii) ; and so by (iii) there must exist a vertex z' such that $xzyx'z'y'x$ is a cycle of length 6 in H_1. But then we will have $f(x) = y'$, $f(y) = x'$ and $f(z') = z'$ and so as $y'z'x'$ is a path in H_1, then $xz'y$ is a path in H_2, therefore $d_{H_2}(x, y) = 2$.

We have proved that if $d_{H_1}(x, y) \leq 2$ then $d_{H_2}(x, y) \leq 2$ and thus $E(H_1^2) \subseteq E(H_2^2)$. Since by construction H_1 and H_2 are isomorphic, H_1^2 and H_2^2 have the same number of edges and we conclude that $H_1 = H_2$. □

Theorem J.11 clearly follows from the proof of Lemma J.12. We end this part with two Lemmas that will turn out to be useful in the next sections. The first one states that if we find a double edge on a cycle of length 6, then the opposite edge is also a double edge ; the second one states that there cannot exist triple edges.

Lemma J.13. *Let H_1, H_2 be distinct coroots in $\mathcal{G}_6$ and $abca'b'c'a$ a cycle of length 6 in H_1, such that $ab \in E(H_1) \cap E(H_2)$. Then $a'b'$ also belongs to $E(H_1) \cap E(H_2)$.*

Proof. Let $D = E(H_1) \cap E(H_2)$ and suppose that $ab \in D$. Then by Lemma J.12 (i), we cannot have $bc \in D$, and by (ii) there cannot be any edge from D incident to c. Thus if we apply (ii) to the edge ca', there must be an edge from D incident to a'. By the same argument, there must be an edge from D adjacent to b' : so we must have $a'b' \in D$, because otherwise there would be two distinct edges from D incident to the edge $a'b'$ respectively in a' and b', which contradicts (ii). □

Lemma J.14. *Let $H_i \in \mathcal{G}_6$, $i \in \{1, 2, 3\}$, be three distinct coroots. Then $E(H_1) \cap E(H_2) \cap E(H_3) = \emptyset$.*

Proof. Suppose that $e = xy \in E(H_1) \cap E(H_2) \cap E(H_3)$, and suppose without loss of generality that y has a neighbour $z \neq x$ in H_1. Then $d_{H_1}(z, x) \leq 2$ and $d_{H_1}(z, y) \leq 2$; so this must also be true in H_2 and H_3, because these three graphs have the same square. So, by Lemma J.3, we must have zx or zy in $E(H_2)$, as well as zx or zy in $E(H_3)$. From this we deduce that at least two of the three graphs share two adjacent edges, which is impossible by Lemma J.12 (i). □

J.5 Two families of graphs

In this section, we define two families of graphs belonging to $\mathcal{G}_6$ and study their coroots. This will prove in particular that (apart from complete graphs) there exist graphs with a number of square roots in $\mathcal{G}_6$ as large as desired.

The graphs $R_{\alpha,\beta}$ and R_α are closely related to the so-called Kneser graphs or bipartite Kneser graphs (e.g. see [21]). For instance, the graph R_α is the generalized Kneser graph $K_{2\alpha-1,\alpha,1}$ and $R_{\alpha,\beta}$ is a generalization of a bipartite Kneser graph. We find it simpler to redefine these graphs and to give them short names closely related to our purpose.

J.5.1 The bipartite graph $R_{\alpha,\beta}$

Let $\alpha \geq 2$ and $\beta \geq 2$ be integers. We define a bipartite graph $R_{\alpha,\beta}$ in the following way :

– the vertex-set of $R_{\alpha,\beta}$ is the disjoint union of the sets $\mathcal{A} = \binom{[\alpha+\beta-1]}{\alpha}$ and $\mathcal{B} = \binom{[\alpha+\beta-1]}{\beta}$. In the case $\alpha = \beta$, let us insist that two disjoint copies of $\binom{[2\alpha-1]}{\alpha}$ are considered ;

– if $A \in \mathcal{A}$ and $B \in \mathcal{B}$, the edge AB is in $E(R_{\alpha,\beta})$ if and only if $|A \cap B| = 1$.

Let us show

Theorem J.15. *For all $\alpha \geq 2$ and $\beta \geq 2$, the graph $R^2_{\alpha,\beta}$ admits exactly $\alpha + \beta$ distinct square roots in $\mathcal{G}_6$.*

Once again it will be convenient to divide the proof into lemmas. Let us denote $[\alpha+\beta-1]$ by N. Recall that for readability we denote set operations such as $A \cup \{a\}$ or $A \setminus B$ respectively by $A + a$ and $A - B$; just beware that in general $A - B + C \neq A + C - B$.

If AB is an edge of $R_{\alpha,\beta}$, we define the *label* of AB to be the element x of N such that $A \cap B = \{x\}$. If $a_1 a_2 \cdots a_k$ is a path in $R_{\alpha,\beta}$, we write

$$a_1 \overset{l_1}{\text{——}} a_2 \overset{l_2}{\text{——}} \cdots \overset{l_{k-1}}{\text{——}} a_k$$

to mean that the label of $a_i a_{i+1}$ is l_i.

Lemma J.16.

(i) *If AB is an edge of $R_{\alpha,\beta}$ with label x, then $B = N - A + x$.*

(ii) *If ABC is a path in $R_{\alpha,\beta}$, where the edges AB and BC are respectively labelled by x and y :*

$$A \overset{x}{\text{——}} B \overset{y}{\text{——}} C,$$

then we have

$$C = A - x + y.$$

Proof. If AB is an edge of $R_{\alpha,\beta}$, with $A \in \mathcal{A}$ and $B \in \mathcal{B}$, then $|A \cap B| = 1$ and

$$\begin{aligned} |A \cup B| &= |A| + |B| - |A \cap B| \\ &= \alpha + \beta - 1 \\ &= |N| , \end{aligned}$$

and so $A \cup B = N$. Since $A \cap B = \{x\}$, (i) follows. Now suppose ABC is a path, as stated in (ii). We have $B = N - A + x$ and $C = N - B + y$, and so

$$C = N - (N - A + x) + y = A - x + y.$$

□

Lemma J.17. *For all $\alpha \geq 2$ and $\beta \geq 2$ the graph $R_{\alpha,\beta}$ has girth* 6.

Proof. Since $R_{\alpha,\beta}$ is bipartite, in order to prove that its girth is at least 6 we just have to show that it contains no cycles of length 4 ; so suppose that $A_1B_1A_2B_2A_1$ is a cycle of length 4. Let us denote the labels of the edges by a_1, b_1, a_2 and b_2 as on Fig. V.

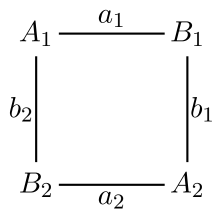


Figure V – *a hypothetical cycle of length four in $R_{\alpha,\beta}$*

By Lemma J.16 we have

$$A_2 = A_1 - a_1 + b_1 \tag{J.2}$$

and

$$A_1 = A_2 - a_2 + b_2. \tag{J.3}$$

and so we must have $\{a_1, a_2\} = \{b_1, b_2\}$. But since $A_2 \neq A_1$ we cannot have $a_1 = b_1$ and since $B_2 \neq B_1$ we can neither have $a_1 = b_2$: a contradiction, therefore the girth of this graph is at least 6.

To conclude, consider $A_0 \in \mathcal{A}$ and suppose without loss of generality that $1 \in A_0$, $3 \in A_0$ but $2 \notin A$. It is easy to see that we can follow from A a single path $A_0A_1A_2A_3A_4A_5A_6$ where the edges are labelled by

$$A_0 \overset{1}{\text{——}} A_1 \overset{2}{\text{——}} A_2 \overset{3}{\text{——}} A_3 \overset{1}{\text{——}} A_4 \overset{2}{\text{——}} A_5 \overset{3}{\text{——}} A_6$$

By Lemma J.16 we have

$$\begin{aligned} A_6 &= A_4 + 3 - 2 \\ &= (A_2 + 1 - 3) + 3 - 2 \\ &= (A_0 + 2 - 1) + 1 - 2 \\ &= A_0 \end{aligned}$$

and so A_0 is on a cycle of length 6. □

Proof of Theorem J.15. Let $x \in N$ and D_x be the set of edges labelled by x in $R_{\alpha,\beta}$. We already know that $R_{\alpha,\beta}$ belongs to $\mathcal{G}_6$; if we show that for all $x \in N$ the set D_x satisfies the conditions of Lemma J.12, then it will prove that $R_{\alpha,\beta}$ admits at least $|N| = \alpha+\beta-1$ coroots in $\mathcal{G}_6$, and so $R^2_{\alpha,\beta}$ will admit at least $\alpha+\beta$ square roots in $\mathcal{G}_6$.

Statement (i) in Lemma J.12 is true for D_x.

This is obvious by Lemma J.16, because if $x \in A$ there is a single edge labelled by x incident to A, which is the edge AB, where

$$B = N - A + x.$$

Statement (ii) in Lemma J.12 is true for D_x.

For similar reasons, if AB is an edge of $R_{\alpha,\beta}$ with label $y \neq x$, then x belongs to exaclty one of the sets A and B ; say A without loss of generality. Then there cannot be edges labelled by x incident to B, and there is a single one in A : the edge AC, where

$$C = N - A + x.$$

Statement (iii) in Lemma J.12 is true for D_x.

Suppose that

$$B_1 \overset{x}{\text{———}} A_1 \overset{a_1}{\text{———}} B_2 \overset{b_2}{\text{———}} A_2 \overset{x}{\text{———}} B_3$$

is a path in $R_{\alpha,\beta}$. Then $a_1 \in B_3$, so if we set $A_3 = N - B_3 + a_1$, the edge B_3A_3 belongs to $E[R_{\alpha,\beta}]$ and is labelled by a_1.

Then by Lemma J.16 we have

$$\begin{aligned} A_3 &= A_1 - a_1 + b_2 - x + a_1 \\ &= A_1 + b_2 - x \\ &= (N - B_1 + x) + b_2 - x \\ &= N - B_1 + b_2, \end{aligned}$$

and so the edge A_3B_1 belongs to $E[R_{\alpha,\beta}]$ and is labelled by b_2. Therefore $B_1A_1B_2A_2B_3A_3B_1$ is a cycle in $R_{\alpha,\beta}$ and (iii) is true.

For all $x \in N$, we have proved that $R_{\alpha,\beta}$ admits a coroot H_x, and all these coroots are obviously distinct since the sets D_x are disjoint by definition of $R_{\alpha,\beta}$. To conclude, suppose that H' is a coroot of H_1. Then by Lemma J.12 (ii), there is at least one edge in $E(H_1) \cap E(H')$. If this edge is labelled by x, from Lemma J.14 we deduce that $H' = H_x$. So $R_{\alpha,\beta}$ admits exactly $|N| = \alpha+\beta-1$ coroots. □

J.5.2 The graph R_α

We define a graph R_α in a similar way for every $\alpha \geq 4$:

– the vertex-set of R_α is $\binom{[2\alpha-1]}{\alpha}$;

– if A and A' belong to $\binom{[2\alpha-1]}{\alpha}$, the edge AA' is in R_α if and only if $|A \cap A'| = 1$.

We define the notion of label of an edge exactly as in $R_{\alpha,\beta}$. It is easy to see that Lemma J.16 still holds in R_α.

As it was done for $R_{\alpha,\beta}$, we can show :

Theorem J.18. *For all* $\alpha \geq 4$, *the graph* R^2_α *admits exactly* 2α *square roots in* $\mathcal{G}_6$.

Lemma J.19. *For all $\alpha \geq 4$ the graph R_α has girth 6.*

Proof. The proof is essentially the same as the previous one, with the difference that R_α is not bipartite. The same arguments than for $R_{\alpha,\beta}$ apply for the inexistence of cycles of length 4 ; therefore the shortest even length of a cycle in R_α is at least 6.

Now we will show that the shortest possible odd length for a cycle in R_α is $2\alpha - 1$. Suppose that

$$A_1 \overset{a_1}{\rule{2em}{0.4pt}} A_2 \overset{a_2}{\rule{2em}{0.4pt}} \cdots A_{2k-1} \overset{a_{2k-1}}{\rule{2em}{0.4pt}} A_1$$

is a cycle of length $2k - 1$ in R_α, where $k \geq 2$.

Then we have

$$A_{2k-1} = A_1 - a_1 + a_2 - a_3 + \cdots - a_{2k-3} + a_{2k-2}$$

and so

$$\begin{aligned} A_1 &= N - A_{2k-1} + a_{2k-1} \\ &= N - A_1 + a_1 - a_2 + a_3 - \cdots + a_{2k-3} - a_{2k-2} + a_{2k-1}. \end{aligned}$$

This implies that $A_1 \subset \{a_1, a_3, \cdots, a_{2k-1}\}$, and so

$$k \geq |A_1| = \alpha.$$

Therefore the length of the cycle is $2k-1 \geq 2\alpha-1 \geq 7$. We have proved that the girth of R_α is at least 6 ; it is easy to show, with the same arguments than for $R_{\alpha,\beta}$, that every vertex of R_α is on a cycle of length 6. Thus R_α has girth 6. □

Proof of Theorem J.18. The proof of Theorem J.15 can be easily adapted to this case. □

J.6 Graphs with multiple roots

We have seen that a complete graph with n vertices admits exactly n square roots in $\mathcal{G}_6$. The following result gives an upper bound for non complete graphs :

Theorem J.20. *A connected non complete graph G with n vertices admits at most*

$$r(n) = \frac{1}{2} + \sqrt{\frac{1}{4} + 2n}$$

square roots in $\mathcal{G}_6$. If $n \geq 11$, then G has exactly $r(n)$ square roots in $\mathcal{G}_6$ if an only if $r(n)$ is an integer and G is isomorphic to the square of $R_{2,r(n)-2}$.

This theorem will be deduced as a corollary to the following one :

Theorem J.21. *Let $H_0 \in \mathcal{G}_6$ be a connected graph admitting exactly r coroots $H_1, H_2, \cdots, H_r$, where $r \geq 3$. Let D be the set of all double edges of the pairs $\{H_0, H_i\}$:*

$$D = \bigcup_{i=1}^{r} E(H_0) \cap E(H_i).$$

Then every connected component of the graph $(V(H_0), D)$ (obtained from H_0 by keeping only double edges) is either :

– *an isolated vertex ;*
– *a star on $r+1$ vertices ;*
– *isomorphic to $R_{\alpha,r+1-\alpha}$ for some integer α such that $2 \leq \alpha \leq r-1$;*
– *isomorphic to $R_{\frac{r+1}{2}}$, this possibility happening only if r is odd and $r \geq 7$.*

Furthermore, if H_0 itself is not a star then one of the two last possibilities happens for at least one component.

J.6.1 Proof of Theorem J.21

It will be convenient to divide the proof in lemmas. We make the following assumptions for this whole section :

> *Hypothesis (∗)*
> Let $H_0 \in \mathcal{G}_6$ a graph admitting exactly exactly r coroots in $\mathcal{G}_6$, where $r \geq 3$, which we denote by $H_1, H_2, \cdots, H_r$.

Now we introduce some notation. Let for each $i \in [r]$

$$D_i = E(H_0 \cap H_i)$$

and

$$D = \bigcup_{i=1}^{r} D_i.$$

If $e \in D_i$, we say that i is the *label* of e ; there is no ambiguity on the labels since the D_i's are disjoint by Lemma J.14. We call D the set of double edges, whereas $E(H_0) \setminus D$ is the set of *unlabelled* edges. If v is a vertex of H_0, we label v by the union $l(v)$ of the labels of the edges incident to v. Thus

$$l(v) = \{i \in [r] \text{ such that } \exists w \in V(H_0),\ vw \in D_i\}$$

is a subset (possibly empty) of $[r]$, and $i \in l(v)$ if and only if there is exactly one edge with label i incident to v.

We will use the labels of vertices to build a correspondence between vertices of H_0 and vertices of the graphs $R_{\alpha,\beta}$, or R_α ; thus it will be convenient to keep our notation from the proof of Theorems J.15 and J.18 for the labels of the edges on a path : a path $v_1 v_2 \cdots v_k$ composed only of double edges where the edge $v_i v_{i+1}$ has label a_i for all $i \in [k-1]$ will be denoted by

$$v_1 \overset{a_1}{\rule{2em}{0.4pt}} v_2 \overset{a_2}{\rule{2em}{0.4pt}} \cdots v_{k-1} \overset{a_{k-1}}{\rule{2em}{0.4pt}} v_k$$

and we will refer to $a_1 a_2 \cdots a_{k-1}$ as the *sequence of labels* of the path $v_1 v_2 \cdots v_k$.

We will use extensively the following lemma during the proof.

Lemma J.22. *With hypothesis (∗), if $x \overset{a}{\rule{2em}{0.4pt}} y$ is an edge labelled by a in H_0, then*

$$l(y) = [r] - l(x) + a.$$

Proof. In the language of labels, Lemma J.12 (i) and (ii) respectively say that adjacent edges cannot bear the same label, and that for every possible label $i \in [r]$, if an edge of H_0 is not labelled by i, then it is adjacent to exactly one edge labelled by i.

So if $x \overset{a}{\rule{2em}{0.4pt}} y$ is an edge in H_0, for every $i \in [r] - a$ such that $i \notin l(x)$, we must have $i \in l(y)$; since we also have $a \in l(y)$, we conclude that $l(y) = [r] - l(x) + a$. □

Let us write two lemmas characterizing the simplest connected components of $(V(H_0), D)$: isolated vertices and stars.

Lemma J.23. *With hypothesis (∗), a vertex $x \in V(H_0)$ is an isolated vertex connected component of $(V(H_0), D)$ if and only if its label is empty.*

Proof. If $l(x) \neq \emptyset$, there must exist a double edge $xy \in D_i$ incident to x for an $i \in \{1, 2, \cdots, r\}$. Thus x and y belong to the same connected component of $(V(H_0), D)$ and x is not isolated.

Conversely, if $l(x) = \emptyset$ then no double edge is incident to x, thus it is isolated in $(V(H_0), D)$. □

Lemma J.24. *With hypothesis (∗), a connected component K of $(V(H_0), D)$ is a star on at least two vertices if and only if it contains a vertex $x \in V(H_0)$ such that $|l(x)| \in \{1, r\}$. In this case it is a star on exactly $r + 1$ vertices.*

Proof. First we show that it is equivalent to require the existence of a vertex x in K with $|l(x)| = 1$ or of a vertex y in K with $|l(y)| = r$. Suppose that $|l(x)| = 1$; $l(x) = \{1\}$ without loss of generality. Then there is a single double edge xy, whose label is 1, incident to x. By Lemma J.22, we have $|l(y)| = r$. Conversely, suppose that K contains a vertex y with $|l(y)| = r$. Then $1 \in l(y)$ and there is a vertex x of K such that the label of xy is 1. Then by Lemma J.22 we have $|l(x)| = 1$.

In both cases, there are exactly r double edges yx_i for $i \in [r]$ incident to y, where yx_i has label i (so $x_1 = x$) ; therefore y has degree r in $(V(H_0), D)$. By Lemma J.22 the vertex x_i has label $l(x_i) = \{i\}$ for every $i \in [r]$, so it has degree 1 in $(V(H_0), D)$. We have proved that K is a star on $r + 1$ vertices with centre y.

Conversely, suppose now that K is a star on at least two vertices. Then it must have a vertex x with degree one in $(V(H_0), D)$, and so $|l(x)| = 1$. □

In Lemmas J.23 and J.24 we have described the nature of the connected components of $(V(H_0), D)$ containing a vertex x with either $|l(x)| = 0$, $|l(x)| = 1$ or $|l(x)| = r$. It remains to study the nature of connected components of $(V(H_0), D)$ where

$$2 \leq |l(x)| \leq r - 1$$

for every vertex x. We call such a connected component *nontrivial*; recall that we want to show that nontrivial connected components of $(V(H_0), D)$ are isomorphic to $R_{\alpha, r+1-\alpha}$ for some $2 \leq \alpha \leq r - 1$ or to $R_{\frac{r+1}{2}}$. First, we show :

Lemma J.25. *With hypothesis (∗), suppose furthermore that H_0 is not a star. Then $(V(H_0), D)$ admits at least one nontrivial connected component.*

Proof. Since $r \geq 3$, by Lemma J.12 the set of double edges D_1 cannot be empty. Let $e = xy \in D_1$. If $2 \leq |l(y)| \leq r - 1$ or $2 \leq |l(x)| \leq r - 1$ then the connected component K of $(V(H_0), D)$ containing xy is nontrivial ; otherwise by Lemmas J.23 and J.24 it must be a star. So let us suppose that K is a star with centre x and leaves $y_1, y_2, \cdots, y_r$, with the edge xy_i labelled by i for every $i \in [r]$.

We have $d_{H_0}(y_2, x) = 1 \leq 2$ and $d_{H_0}(y_2, y_1) = 2$; but H_0 and H_1 are coroots, therefore we must also have $d_{H_1}(y_2, x) \leq 2$ and $d_{H_1}(y_2, y_1) \leq 2$. Since we assumed that $xy_1 \in E(H_1)$, by the Adjacency Lemma J.3 we must have $y_2y_1 \in E(H_1)$, and so y_1 has degree at least 2 in H_1 (in fact its degree is at least r since we could have shown $y_iy_1 \in E(H_1)$ for every $i > 1$). We now use the fact that H_0 is not a star : by Lemma J.6 if a vertex of H_0 has degree one in H_0 then it must have degree one in every coroot of H_0. Since we proved that the degree of y_1 is at least 2 in H_1, we deduce that the degree of y_1 must be at least 2 in H_0.

So there must exist an edge $y_1 z$ incident to y_1 with $z \neq x$. Since $l(y_1) = \{1\}$, this edge is necessarily unlabelled (see fig. VI).

By Lemma J.12, the labels of an unlabelled edge must be complementary sets in $[r]$, so we must have

$$|l(z)| = r - 1,$$

and so the connected component of $(V(H_0), D)$ containing z must be nontrivial. □

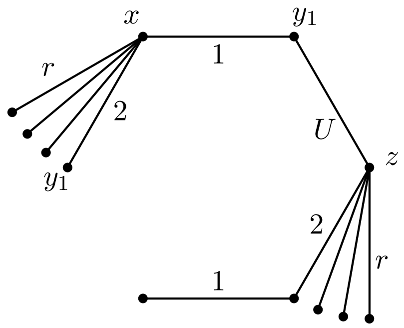


Figure VI – *proof of Lemma J.25.*

Lemma J.26. *With hypothesis (∗), if*

$$v_1 \overset{a_1}{\text{——}} v_2 \overset{a_2}{\text{——}} \cdots v_{2k} \overset{a_{2k}}{\text{——}} v_{2k+1}$$

is a walk of double edges with even length, where $k \geq 1$, then

$$l(v_{2k+1}) = l(v_1) - a_1 + a_2 - a_3 + a_4 + \cdots - a_{2k-1} + a_{2k}.$$

Proof. Recall that in a walk vertices are not necessarily distinct. By Lemma J.22, one has

$$\begin{aligned} l(v_3) &= [r] - l(v_2) + a_2 \\ &= [r] - ([r] - l(v_1) + a_1) + a_2 \\ &= l(v_1) - a_1 + a_2, \end{aligned}$$

and then the result is obvious by induction. □

Lemma J.27. *With hypothesis (∗), let K be a nontrivial connected component of $(V(H_0), D)$, and let $xy \in E(K)$. Let $\alpha = |l(x)|$ and $\beta = |l(y)|$. Then every vertex $v \in W$ satisfies $|l(v)| \in \{\alpha, \beta\}$ and conversely, for every $A \in \binom{[\alpha+\beta-1]}{\alpha} \cup \binom{[\alpha+\beta-1]}{\beta}$, there is a vertex of K with label A.*

Proof. By Lemma J.22, if vw is a double edge, we have

$$|l(v)| + |l(w)| = r + 1.$$

Since $\alpha + \beta = r + 1$, one has $|l(v)| = \alpha$ if and only if $|l(w)| = \beta$; thus the first part of the Lemma is clear since every pair of vertices can be linked by a path of double edges in K.

Consider now a vertex v_0 with label A and choose $a \in A$; there exists v_1 such that $v_0 v_1$ is labelled by a. As

$$l(v_1) = [r] - l(v_0) + a,$$

if we choose $b \notin l(v_0)$, there is a vertex v_2 such that v_1v_2 is labelled by b. So we have

$$l(v_2) = l(v_0) - a + b = A - a + b.$$

We just showed that for every $a \in A$ and $b \notin A$, there is in K a vertex with label $A - a + b$. This clearly implies the second part of the lemma by induction. □

We have just shown that the labels of vertices in a nontrivial component K of $(V(H_0), D)$ are all the possible subsets of $\binom{[\alpha+\beta-1]}{\alpha}$ and of $\binom{[\alpha+\beta-1]}{\beta}$, but we do not know yet how many times a given label can appear in $V(K)$. The next lemma deals with the simplest case.

Lemma J.28. *With hypothesis* $(*)$, *let* K *be a nontrivial connected component of* $(V(H_0), D)$ *and suppose furthermore that there are no distinct vertices sharing the same label in* $V(K)$. *Then either* $\alpha \neq \beta$ *and* K *is isomorphic to* $R_{\alpha,\beta}$, *or* r *is odd,* $\alpha = \beta = \frac{r+1}{2} \geq 4$, *and* K *is isomorphic to* R_α.

Proof. Let us define $R = R_{\alpha,\beta}$ if $\alpha \neq \beta$ and $R = R_\alpha$ if $\alpha = \beta$. By Lemma J.27 and the assumption that distinct vertices of K have distinct labels, the map l which assigns its label to every vertex is a bijection from $V(K)$ to $\binom{[\alpha+\beta-1]}{\alpha} \cup \binom{[\alpha+\beta-1]}{\beta}$ in case of $\alpha \neq \beta$, or to $\binom{[2\alpha-1]}{\alpha}$ in case of $\alpha = \beta$. Thus l defines a bijection map from $V(K)$ to the vertex set of R.

Suppose that v, w are adjacent vertices in K. Then there is i such that $vw \in D_i$, and so

$$l(v) \cap l(w) = \{i\}.$$

Thus $l(v)$ and $l(w)$, considered as vertices of R, are adjacent.

Conversely, let us consider two adjacent vertices in R; we may denote them by $l(v)$ and $l(w)$, where $v, w \in V(K)$, and there is i such that $l(v) \cap l(w) = \{i\}$. We have $i \in l(v)$, so v admits a neighbour z, such that $vz \in D_i$, therefore

$$l(z) = [r] - l(v) + i = l(w),$$

and so $z = w$ by uniqueness of the label in K : thus v and w are adjacent in K. We have shown that l is an isomorphism from K to R. In the case $\alpha = \beta$, since $\alpha + \beta = r + 1$, we see that $\alpha = \frac{r+1}{2}$; since R_3 has girth 5 (e.g. consider the vertex $\{1; 3; 5\}$ and follow the path with sequence of labels 12345), we must have $\alpha \geq 4$. □

It remains to study the case where labels appear several times in K, which is more involved. We will show that if one label appears at least twice in K, then every label appears exactly twice. Next, we prove that in this case K is isomorphic to $R_{\alpha,\alpha}$. We need the following notation : for all $k \geq 3$, let us denote by $\Sigma'(k)$ the subgroup of the permutation group of $[k]$ generated by the transpositions $(i, i+2)$ for every $i \in [k-2]$. The group $\Sigma'(k)$ is precisely the set of permutations of $[k]$ which map even integers onto even integers (and odd integers onto odd integers), i.e. $\Sigma'(k)$ is the set of all bijective functions σ from $[k]$ to itself, such that

$$\forall i \in [k],\ \sigma(i) \equiv i\ [2].$$

Let us also define a *shortest path of double edges* between two vertices v, w to be a path consisting only of double edges from v to w, with the smallest possible length. By definition of K, there is a shortest path of double edges between any two vertices of K.

We begin this part by a technical lemma.

Lemma J.29. *With hypothesis $(*)$, let K be a nontrivial connected component of $(V(H_0), D)$. Let $k \geq 3$ and*

$$v_1 \overset{a_1}{\rule{2em}{0.4pt}} v_2 \overset{a_2}{\rule{2em}{0.4pt}} \cdots v_k \overset{a_k}{\rule{2em}{0.4pt}} v_{k+1}$$

be a shortest path of double edges between two vertices v_1 and v_{k+1} of K. Then :

(i) the labels of this path are distinct : $a_i \neq a_j$ whenever $i \neq j$;

(ii) If $a'_1 a'_2 \cdots a'_k$ is permutation of the sequence $a_1 a_2 \cdots a_k$ by an element of $\Sigma'(k)$, then there is a shortest path of double edges from v_1 to v_{k+1} with labels

$$v_1 \overset{a'_1}{\rule{2em}{0.4pt}} \cdot \overset{a'_2}{\rule{2em}{0.4pt}} \cdots \cdot \overset{a'_k}{\rule{2em}{0.4pt}} v_{k+1}$$

(iii) Conversely, if

$$v_1 \overset{a'_1}{\rule{2em}{0.4pt}} v'_2 \overset{a'_2}{\rule{2em}{0.4pt}} \cdots v'_k \overset{a'_k}{\rule{2em}{0.4pt}} v'_{k+1}$$

is a shortest path of double edges from v_1 to a vertex v'_{k+1} and the sequence $a'_1 a'_2 \cdots a'_k$ is a permutation of $a_1 a_2 \cdots a_k$ by an element of $\Sigma'(k)$, then $v'_{k+1} = v_{k+1}$.

Proof. Suppose that there is a shortest path of double edges from v_1 to v_{k+1} where two labels a_i and a_j are equal for $1 \leq i < j \leq k$, and choose this path such that (i, j) is minimal in the lexicographic order.

First note that $j \geq i + 3$, because by Lemma J.12 two adjacent edges cannot bear the same label, and if $j = i + 2$

$$v_i \overset{a_i}{\rule{2em}{0.4pt}} v_{i+1} \rule{2em}{0.4pt} v_{i+2} \overset{a_i}{\rule{2em}{0.4pt}} v_{i+3}$$

the edge $v_{i+1}v_{i+2}$ would be adjacent to two edges labelled by a_i, contradicting Lemma J.12 once again. So we must have $j \geq 4$.

Let us consider the path

$$v_{j-2} \overset{a_{j-2}}{\rule{2em}{0.4pt}} v_{j-1} \overset{a_{j-1}}{\rule{2em}{0.4pt}} v_j \overset{a_j}{\rule{2em}{0.4pt}} v_{j+1}.$$

We have $a_j \notin l(v_{j-1})$ and so $a_j \in l(v_{j-2})$. Thus there is a vertex v'_{j-1} in H_0 such that $v_{j-2}v'_{j-1} \in D_{a_i}$; clearly v'_{j-1} must be distinct from v_{j-2}, v_{j-1}, v_j and v_{j+1}. Then we can apply Lemma J.12 (iii) to the path $v'_{j-1}v_{j-2}v_{j-1}v_jv_{j+1}$: there is a vertex v'_j such that

$$v_{j-2}\ v_{j-1}\ v_j\ v_{j+1}\ v'_j\ v'_{j-1}\ v_{j-1}$$

is a cycle of length 6 in H_0. We cannot have $v_{j-3} = v'_{j-1}$ otherwise it would contradict the minimality of length of the path of double edges. By Lemma J.13 we infer that the respective labels of $v'_{j-1}v'_j$ and v'_jv_{j+1} are a_{j-1} and a_{j-2}. The situation is depicted on Fig. VII.

Using this cycle, in the original path from v_1 to v_{k+1}, we can replace the path

$$v_{j-2} \overset{a_{j-2}}{\rule{2em}{0.4pt}} v_{j-1} \overset{a_{j-1}}{\rule{2em}{0.4pt}} v_j \overset{a_j}{\rule{2em}{0.4pt}} v_{j+1}.$$

by

$$v_{j-2} \overset{a_j}{\rule{2em}{0.4pt}} v'_{j-1} \overset{a_{j-1}}{\rule{2em}{0.4pt}} v'_j \overset{a_{j-2}}{\rule{2em}{0.4pt}} v_{j+1}.$$

Doing so, we obtain a new shortest path of double edges from v_1 to v_{k+1}. In this path the edges with position i and $j - 2$ share the same label, and since $(i, j - 2)$ comes before (i, j) in the lexicographic order we have a contradiction. Thus (i) is true.

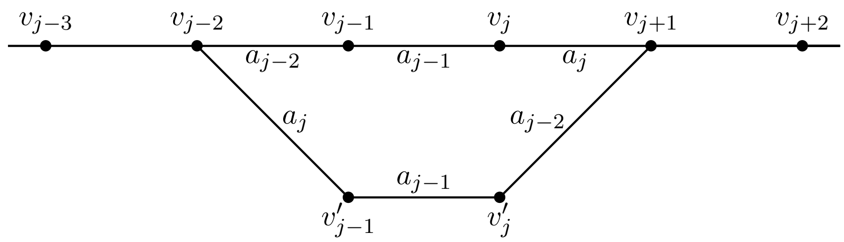


Figure VII – *Proof of Lemma J.29*

Consider now once again a shortest path of double edges

$$v_1 \overset{a_1}{\text{———}} \cdot \overset{a_2}{\text{———}} \cdots \cdot \overset{a_k}{\text{———}} v_{k+1}.$$

Since all the labels on this path are distinct, for every $i \in [k-2]$ we can apply the technique used for (i) in order to obtain a shortest path of double edges from v_1 to v_{k+1}, where the labels are the same apart from the fact that a_i and a_{i+2} have been exchanged. Since we defined $\Sigma'(k)$ as the subgroup of the permutation group of $[k]$ generated by the transpositions $(i, i+2)$, we see that by doing this repeatedly we can obtain any shortest path of double edges from v_1 to v_{k+1} whose sequence of labels is obtained from $a_1 a_2 \cdots a_k$ by an element of $\Sigma'(k)$. Therefore (ii) is true.

By (ii) and since at most one path with a given sequence of labels can be issued from a given vertex, (iii) easily follows. □

Lemma J.30. *With hypothesis* $(*)$*, let* K *be a connected component of* $(V(H_0), D)$ *and suppose that there are two distinct vertices* v, v' *of* K *sharing the same label. Then*

(i) $\alpha = \beta$ *;*

(ii) *each possible label* $A \in \binom{[2\alpha-1]}{\alpha}$ *appears exactly twice in* W *;*

(iii) *any shortest path of double edges from* v *to* v' *has length* $2\alpha - 1$ *and if we denote its labels by*

$$v = v_1 \overset{a_1}{\text{———}} v_2 \overset{a_2}{\text{———}} \cdots v_{2\alpha-1} \overset{a_{2\alpha-1}}{\text{———}} v_{2\alpha} = v'$$

then we have

$$l(v) = \{a_1, a_3, \cdots, a_{2\alpha-1}\} \tag{J.4}$$

and

$$[r] - l(v) = \{a_2, a_, \cdots, a_{2\alpha-2}\}; \tag{J.5}$$

(iv) *conversely, for every sequence of distinct labels* $a_1 a_2 \cdots a_{2\alpha-1}$ *such that (J.4) and (J.5) are satisfied, there is a shortest path of double edges from* v *to* v' *with this sequence of labels.*

Proof. Suppose that v and v' are distinct vertices in W such hat $l(v) = l(v')$. By definition of W there is a shortest path of double edges

$$v = v_1 \overset{a_1}{\text{———}} v_2 \overset{a_2}{\text{———}} \cdots v_k \overset{a_k}{\text{———}} v_{k+1} = v'$$

from v to v'. The labels on this path are distinct by Lemma J.29; so if k is even, by Lemma J.26 we have

$$l(v') = l(v) - (a_1 + a_3 + \cdots + a_{k-1}) + (a_2 + a_4 + \cdots + a_k)$$

and so

$$\{a_1, a_3, \cdots, a_{k-1}\} = \{a_2, a_4, \cdots, a_k\},$$

which is impossible; so k is odd. By Lemma J.22 we know that along a path of double edges in W the labels have alternatively cardinality α and β; since $l(v_1) = l(v_{k+1})$ and k is odd then $\alpha = \beta$, which proves (*i*).

Since $k-1$ is even one has by Lemma J.26

$$l(v_k) = l(v_1) - (a_1 + a_3 + \cdots + a_{k-2}) + (a_2 + a_4 + \cdots + a_{k-1})$$

and so

$$\begin{aligned} l(v_{k+1}) &= [r] - l(v_k) + a_k \\ &= [r] - l(v_1) + (a_1 + a_3 + \cdots + a_{k-2}) + a_k - (a_2 + a_4 + \cdots + a_{k-1}). \end{aligned}$$

Since $l(v_{k+1}) = l(v_k)$ and the a_i's are distinct, we must have

$$l(v_1) = \{a_1, a_3, \cdots, +a_k\} \tag{J.6}$$

and

$$[r] - l(v_1) = \{a_2, a_4, \cdots, a_{k-1}\}. \tag{J.7}$$

Moreover, since $l(v_1)$ has size α, we see that $\frac{k+1}{2} = \alpha$, thus the length of the path is $k = 2\alpha - 1$. We have proved (*iii*).

Suppose now that another vertex v'_{k+1} shares the label of v and v' and consider a shortest path of double edges from $v = v_1$ to v'_{k+1} with sequence of labels $a'_1 a'_2 \cdots a'_l$; then we must have $l = 2\alpha - 1$, so (J.6) and (J.7) must also hold for the a'_i's. Therefore, the sequence of labels $a'_1 a'_2 \cdots a'_{2\alpha-1}$ can be obtained from $a_1 a_2 \cdots a_{2\alpha-1}$ by a permutation which belongs to $\Sigma'(2\alpha - 1)$. By Lemma J.29 (*iii*), we deduce that $v'_{k+1} = v_{k+1}$. Thus at most two vertices of K can share the same label.

Let us show now, that if we suppose that at least one label appears twice in K, then every possible label $A \in \binom{[2\alpha-1]}{\alpha}$ appears twice in K. Keeping the same notation, let us consider v_1 and $v_{2\alpha}$, two distinct vertices of K with the same label. Consider now a neighbour w_1 of v_1 such that $v_1 w_1 \in D_i$ for some $i \in [r]$. Then there is a neighbour $w_{2\alpha}$ of $v_{2\alpha}$ such that $v_{2\alpha} w_{2\alpha} \in D_i$. Clearly, we have $w_1 \neq w_{2\alpha}$, and since $l(v_1) = l(v_{2\alpha})$ we have

$$l(w_1) = [r] - l(v_1) + i = [r] - l(v_{2\alpha}) + i = l(w_{2\alpha}).$$

So the label of w_1 also appears at least twice in K. By a simple argument of connectivity, we see the same conclusion is true for every vertex in K. This proves (*ii*).

It remains to see (*iv*), which is a consequence of Lemma J.29 (*iii*). □

With the same hypothesis, we go further in the next lemma.

Lemma J.31. *With hypothesis (∗), let K be a connected component of $(V(H_0), D)$ and suppose that there are two distinct vertices v, v' of K sharing the same label. Then K is bipartite.*

Proof. We have to show that there cannot exist a cycle of double edges with odd length in H_0, so suppose that

$$C: \; v_1 \overset{a_1}{\text{——}} v_2 \overset{a_2}{\text{——}} \cdots v_{2k-1} \overset{a_{2k-1}}{\text{——}} v_{2k} = v_1$$

is one, with the minimum possible length.

Since $2k-1$ is odd we have

$$l(v_1) = [r] - l(v_1) + a_1 - a_2 + a_3 + \cdots - a_{2k-2} + a_{2k-1},$$

so if we show that all the labels a_i are distinct on the cycle, it will follow that $k = \alpha$ and

$$l(v_1) = \{a_1, a_3, \cdots, a_{2k-1}\} \tag{J.8}$$

$$[r] - l(v_1) = \{a_2, a_4, \cdots, a_{2k-2}\}, \tag{J.9}$$

a contradiction with Lemma J.30 where it is proved that if we follow, starting from v_1 that a path with sequence of labels $a_1a_2 \cdots a_{2\alpha-1}$ satisfying (J.8) and (J.9), we arrive at a vertex v' with $l(v') = l(v_1)$ but $v' \neq v_1$.

Thus, we have to show that all the a_i's are distinct on the cycle

$$C \ : \ v_1 \overset{a_1}{\overline{\quad}} v_2 \overset{a_2}{\overline{\quad}} \cdots v_{2k-1} \overset{a_{2k-1}}{\overline{\quad}} v_{2k} = v_1.$$

Let us notice first that the minimality of C implies that there cannot exist a double edge between two non consecutive vertices of C. Indeed, suppose without loss of generality that v_1v_i is a double edge where $2 < i < 2k-1$. The edge v_1v_i cuts the cycle in two paths which are $p_1 = v_1v_2 \cdots v_i$ and $p_2 = v_iv_{i+1} \cdots v_{2k-1}v_1$; since the length of the cycle is odd, p_1 or p_2 must have even length. In both cases this path of even length, followed by v_1v_i, yields to a cycle of double edges with odd length which contradicts the minimality of C.

Let us show, as in the proof of Lemma J.30, that for every i there is a cycle whose sequence of labels is obtained just by permuting a_i and a_{i+2} in the sequence (here $i+2$ is understood modulo $2k-1$).

Without loss of generality we prove this for $i = 1$. Let us now consider the three consecutive labels

$$v_1 \overset{a_1}{\overline{\quad}} v_2 \overset{a_2}{\overline{\quad}} v_3 \overset{a_3}{\overline{\quad}} v_4.$$

By Lemma J.12 we know that a_1, a_2 and a_3 must be distinct and as

$$l(v_4) = l(v_2) - a_2 + a_3,$$

we have $a_1 \in l(v_4)$; so there must exist a vertex x such that $v_4x \in D_{a_1}$. Then by Lemma J.12 (*iii*) there must exist a vertex y such that $v_1v_2v_3v_4xyv_1$ is a cycle of double edges of length 6. Let us show that x and y cannot belong to C.

First, as we observed that two vertices of C adjacent by a double edge must be consecutive, we have either $x \notin C$ or $x = v_5$.

- If $x = v_5$, then either
  - $y = v_6$, and in this case the edge v_6v_1 would belong to C, which is impossible ;
  - or $y \neq v_6$, but in that case $v_1yv_5v_6 \cdots v_{2k-1}v_1$ is a cycle of double edges with odd length, contradicting the minimality of C ;
- and if $x \neq v_5$, we cannot have $y = v_{2k-1}$ which once again would create a cycle of double edges with a smaller odd length.

Thus we conclude that x and y do not belong to C, and in the cycle we can replace the path $v_1v_2v_3v_4$ by the path v_1yxv_4 , and obtain a cycle of double edges

$$C' \ v_1 \overset{a_3}{\overline{\quad}} y \overset{a_2}{\overline{\quad}} x \overset{a_1}{\overline{\quad}} v_4 \overset{a_1}{\overline{\quad}} v_5 \cdots v_{2k-1} \overset{a_{2k-1}}{\overline{\quad}} v_{2k} = v_1$$

which has the same length than C and the same sequence of labels, apart from the fact that a_1 and a_3 have been exchanged.

Since the length of C is odd, we see that we can obtain the existence of a cycle of double edges of length $2k-1$ with labels $a'_1 a'_2 \cdots a'_{2k-1}$ for any permutation of the original sequence $a_1 a_2 \cdots a_{2k-1}$. This implies that a label can appear at most once on the cycle, because otherwise we could obtain by permutation adjacent edges bearing the same label, which is impossible by Lemma J.12. This ends the proof of this lemma. □

The next lemma is the last one needed for the proof of Theorem J.21.

Lemma J.32. *With hypothesis (∗), let K be a connected component of $(V(H_0), D)$ and suppose that there are two distinct vertices v, v' of K sharing the same label. Then r is odd and K is isomorphic to $R_{\alpha,\alpha}$ with $\alpha = \frac{r+1}{2}$.*

Proof. The graph $R_{\alpha,\alpha}$ is bipartite ; let $R_1 \cup R_2$ be its bipartition, where R_1 and R_2 are copies of $\binom{[2\alpha-1]}{\alpha}$. We just proved in Lemma J.31 that under these assumptions the subgraph K of H_1 must be bipartite ; let $V(K) = W_1 \cup W_2$ be its bipartition. We have also proved in Lemma J.30 that if $v \neq v' \in V(K)$ share the same label then the length of any shortest path of double edges from v to v' is even ; so v and v' must lie on different sides of the bipartition. Since each possible label $A \in \binom{[2\alpha-1]}{\alpha}$ appears exactly twice in K, we see that the map which assigns its label to every vertex defines a bijection from $V(K)$ the vertex set of $R_{\alpha,\alpha}$, mapping W_1 onto R_1 and W_2 onto R_2. It remains to prove that this map is an isomorphism, and this can be done by following straightforwardly the proof of Lemma J.28. □

With the help of the different lemmas we can prove Theorem J.21.

Proof of Theorem J.21.

Let K be a connected component of $(V(H_0), D)$. If there is a vertex x in K with $|l(x)| \in \{0, 1, r\}$, then by Lemmas J.23 and J.24 K is either an isolated vertex or a star on $r+1$ vertices ; and if such an x doesn't exist K is either isomorphic to $R_{\alpha,\beta}$ for $\alpha \geq 2$, $\beta \geq 2$ and $\alpha + \beta = r + 1$ (Lemma J.28), or to $R_{\frac{r+1}{2}}$ (Lemma J.32). We have also proved in Lemma J.25 that if H_0 is not a star then $(V(H_0), D)$ admits at least one nontrivial connected component. Therefore, Theorem J.21 follows. □

J.6.2 Proof of Theorem J.20

Proof of Theorem J.20. Let G be a connected graph, admitting exactly r square roots in $\mathcal{G}_6$, which is not complete. If G has radius one, then $r \leq 2$ by Lemma J.8, and since $r(n) \geq 2$ for $n \geq 1$ we have $r \leq r(n)$. Suppose now that G has radius at least 2. By Theorem J.11 we know that the square roots in $\mathcal{G}_6$ of G are all isomorphic, and they must contain a cycle of length 6 by Theorem J.9. Therefore the number n of vertices of G is at least 6. Since

$$\frac{1}{2} + \sqrt{\frac{1}{4} + 2 \times 6} > 3,$$

our bound is satisfied in the case $r \leq 3$, and there cannot be equality. We suppose from now on that $r \geq 4$.

Let $H_0 \in \mathcal{G}_6$ be a square root of G ; it satisfies the hypothesis of Theorem J.21 with $r-1$ coroots and thus H_0 contains a subgraph K isomorphic either to $R_{\alpha,r-\alpha}$ for some

$2 \leq \alpha \leq r-2$ or (for even r with $r > 4$ to $R_{\frac{r}{2}}$. It is easy to see that the smallest of these graphs is $R_{2,r-2}$ and that its number of vertices is $\binom{r}{2}$. We conclude that

$$n \geq \frac{r(r-1)}{2}$$

and so

$$r \leq \frac{1}{2} + \sqrt{\frac{1}{4} + 2n}$$

by solving this inequality. There can be equality only if K is isomorphic to $R_{2,r-2}$ and is a spanning subgraph of H_0, i.e. K can be obtained from H_0 only by deletion of edges. Let us suppose that there is an edge $e = xy \in E(H_0) - E(K)$. Since the edge xy cannot be labelled (otherwise it would belong to $E(K)$) we must have

$$l(x) \cap l(y) = \emptyset$$

and by Lemma J.12 (*ii*)

$$l(x) \cup l(y) = [r-1]$$

and so $|l(x)| + |l(y)| = r - 1$. Since $|l(x)|$ and $|l(y)|$ must be equal to 2 or $r-2$, we see that the only possibilities are $|l(x)| = 2$ and $|l(y)| = 2$, which leads to $r = 5$, and $|l(x)| = r-2$ and $|l(y)| = r-2$, which leads to $r = 3$. The number of vertices corresponding to these two cases are respectively 10 and 6, so if $n \geq 11$ and the bound is tight, the spanning subgraph K of H_0 is an induced subgraph of H_0 and thus equal to H_0. ☐

We proved that the graph $R_{2,r-2}$ has the maximum number of coroots in $\mathcal{G}_6$ for infinitely many values of n. This graph is easy to describe : it can be constructed by subdividing every edge of the complete graph K_{r-1} exactly once. For $r = 4$, we obtain the cycle C_6 ; for $r = 5$ we obtain a graph with 4 coroots, which is depicted in Figure VIII.

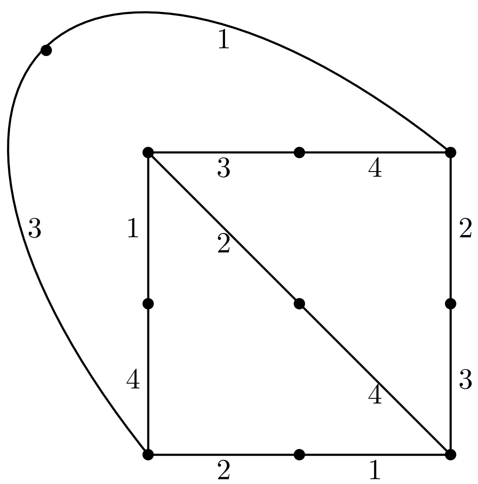


Figure VIII – *The graph $R_{2,4}$ admits 4 coroots. We have figured the labels of double edges on the graph.*

J.7 Remarks and open questions

• We have seen that the problem Π_k :

Given a connected graph G, is there a graph $H \in \mathcal{G}_k$ such that $H^2 = G$?

is polynomial for $k \geq 6$, whereas it is NP-complete in the general case $k = 3$. We are naturally interested in the algorithmic complexity of Π_k for $k = 4$ or $k = 5$. We tend to believe that these problems are NP-complete. One of our reasons is the following observation : there is a family of graphs in $\mathcal{G}_5$ with an exponential number of coroots. A member of this family is depicted in Figure IX. It is built with a central star with a variable number of leaves (four in Fig. IX) ; to each leaf is attached a cycle on 5 vertices, either twisted or non twisted. It is easy to check that the various graphs obtained for every choice of the cycles have the same square.

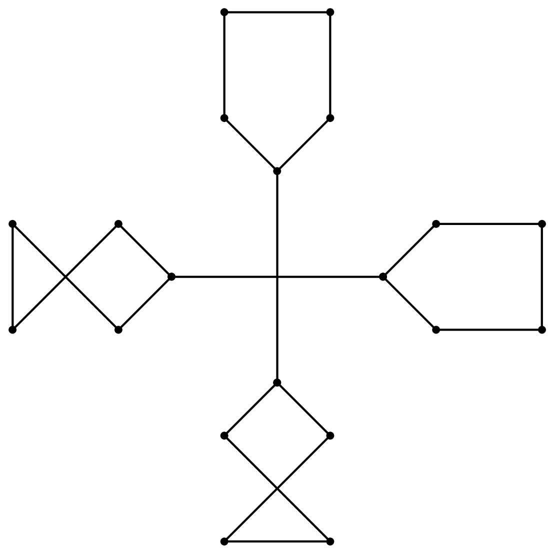


Figure IX – *A graph that admits* 15 *coroots in* $\mathcal{G}_5$, *obtained by twisting or not twisting the cycles of length* 5.

However, an exponential number of square roots in not necessarily an obtruction to the polynomiality of Π_5 ; for instance, Lau ([71]) proved that the square of a bipartite graph may have an exponential number of square roots, but that the problem of recognition of these graphs is polynomial.

• We have shown that graphs admitting multiple roots in $\mathcal{G}_6$ have a very regular structure, and that their different square roots can be deduced one from each other by twisting opposite edges on cycles with length 6. We would be interested in similar results for graphs having multiple roots in $\mathcal{G}_5$. Are their square roots isomorphic ? Do these graphs admit a particular structure ?

• Theorem J.11 has as a consequence that a graph cannot admit simultaneously a square root of girth 6 and a square root with higher girth. However, is it possible for a graph to admit a square root with girth at least 6 and another with girth 5 or less ? This is surely the case for complete graphs.

• Finally, we believe that similar results could be obtained for powers H^r of a graph H with $r \geq 3$. For instance, it seems plausible to obtain the unicity of the r-th root of a graph with sufficient diameter in $\mathcal{G}_k$ for k sufficiently high.

Bibliographie

[1] D. Aingworth, R. Motwani, and F. Harary. The difference between a graph and its square. *Utilitas Mathematica*, pages 223–228, 1998.

[2] S. Arnborg, J. Lagergren, and D. Seese. Easy problems for tree-decomposable graphs. *J. Algorithms*, 12(2) :308–340, 1991.

[3] D. Auger. Minimal identifying codes in trees and planar graphs with large girth. *European Journal of Combinatorics, à paraître.*

[4] D. Auger. Problèmes d'identification métrique dans les graphes. *Rapport interne Télécom Paris-2007D013, Paris, France.*

[5] D. Auger. Induced paths in twin-free graphs. *Electronic Journal of Combinatorics*, 15(1)(17), 2008.

[6] D. Auger, I. Charon, I. Honkala, O. Hudry, and A. Lobstein. Edge number, minimum degree, maximum independent set, radius and diameter in twin-free graphs. *Advances in Mathematics of Communications*, 3(1) :97–114, 2009.

[7] D. Auger, I. Charon, O. Hudry, and A. Lobstein. Complexity results for identifying codes in planar graphs. *International Transactions in Operational Research, à paraître.*

[8] D. Auger, I. Charon, O. Hudry, and A. Lobstein. Maximum size of a minimum watching system and the graphs achieving the bound. *Soumis pour publication.*

[9] D. Auger, I. Charon, O. Hudry, and A. Lobstein. On the existence of a cycle of length at least 7 in a $(1, \leq 2)$-twin-free graph. *Discussiones Mathematicae-Graph Theory, to appear.*

[10] D. Auger, I. Charon, O. Hudry, and A. Lobstein. On the sizes of the graphs G, G^r, $G^r \setminus G$: the directed case. *Soumis pour publication.*

[11] D. Auger, I. Charon, O. Hudry, and A. Lobstein. On the sizes of the graphs G, G^r, $G^r \setminus G$: the undirected case. *Soumis pour publication.*

[12] D. Auger, I. Charon, O. Hudry, and A. Lobstein. On the square roots of graphs. *Soumis pour publication.*

[13] D. Auger, I. Charon, O. Hudry, and A. Lobstein. Watching systems in graphs : an extension of identifying codes. *Soumis pour publication.*

[14] D. Auger, I. Charon, O. Hudry, and A. Lobstein. Existence d'un cycle de longueur au moins 7 dans un graphe sans $(1, \leq 2)$-jumeaux. *Rapport interne Telecom ParisTech-2009D015, Paris, France*, 2009.

[15] J.P. Barthélemy, G. Cohen, A. Lobstein, and M. Minoux. *Complexité algorithmique et problèmes de communications.* Masson, 1992.

[16] C. Berge. *Graphes*. Gauthier-Villars Paris, 1983.

[17] N. Bertrand, I. Charon, O. Hudry, and A. Lobstein. Identifying and locating-dominating codes on chains and cycles. *European Journal of Combinatorics*, 25(7) :969–987, 2004.

[18] N. Bertrand, I. Charon, O. Hudry, and A. Lobstein. 1-identifying codes on trees. *Australasian Journal of Combinatorics*, 31 :21–35, 2005.

[19] M. Blidia, M. Chellali, F. Maffray, J. Moncel, and A. Semri. Locating-domination and identifying codes in trees. *Australasian Journal of Combinatorics*, 39 :219, 2007.

[20] JA Bondy. Induced subsets. *Journal of Combinatorial Theory, Series B*, 12(2) :201–202, 1972.

[21] J.A. Bondy and U.S.R. Murty. *Graph Theory, volume 244 of Graduate Texts in Mathematics*. Springer, New York, 2008.

[22] F. Buckley and F. Harary. *Distance in graphs*. Addison-Wesley Reading, MA, 1990.

[23] E. Charbit, I. Charon, G. Cohen, and O. Hudry. Discriminating codes in bipartite graphs. *Electronic Notes in Discrete Mathematics*, 26 :29–35, 2006.

[24] E. Charbit, I. Charon, G. Cohen, O. Hudry, and A. Lobstein. Discriminating codes in bipartite graphs : bounds, extremal cardinalities, complexity. *Advances in Mathematics of Communications*, 4(2) :403–420, 2008.

[25] I. Charon, G. Cohen, O. Hudry, and A. Lobstein. Links between discriminating and identifying codes in the binary Hamming space. *Lecture Notes in Computer Science*, 4851 :267, 2007.

[26] I. Charon, G. Cohen, O. Hudry, and A. Lobstein. Discriminating codes in (bipartite) planar graphs. *European Journal of Combinatorics*, 29(5) :1353–1364, 2008.

[27] I. Charon, S. Gravier, O. Hudry, A. Lobstein, M. Mollard, and J. Moncel. A linear algorithm for minimum 1-identifying codes in oriented trees. *Discrete Applied Mathematics*, 154(8) :1246–1253, 2006.

[28] I. Charon, I. Honkala, O. Hudry, and A. Lobstein. The minimum density of an identifying code in the king lattice. *Discrete Mathematics*, 276(1-3) :95–109, 2004.

[29] I. Charon, I. Honkala, O. Hudry, and A. Lobstein. Structural properties of twin-free graphs. *Electronic Journal of Combinatorics*, 14, 2007.

[30] I. Charon, O. Hudry, and A. Lobstein. Extremal values for identification, domination and maximum cliques in twin-free graphs. *Ars Combinatoria, à paraître*.

[31] I. Charon, O. Hudry, and A. Lobstein. Extremal values for the maximum degree in a twin-free graph. *Ars Combinatoria, à paraître*.

[32] I. Charon, O. Hudry, and A. Lobstein. Identifying codes with small radius in some infinite regular graphs. *Electronic Journal of Combionatorics*, 9(1), 2002.

[33] I. Charon, O. Hudry, and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Rapport interne Télécom Paris-2003D006, Paris, France*, page 18, 2003.

[34] I. Charon, O. Hudry, and A. Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoretical Computer Science*, 290(3) :2109–2120, 2003.

[35] I. Charon, O. Hudry, and A. Lobstein. On the structure of identifiable graphs : results, conjectures, and open problems. *Proceedings 29th Australasian Conference in Combinatorial Mathematics and Combinatorial Computing,Taupo, New Zealand*, pages 37–38, 2004.

[36] I. Charon, O. Hudry, and A. Lobstein. On the structure of identifiable graphs. *Electronic Notes in Discrete Mathematics*, 22 :491–495, 2005.

[37] I. Charon, O. Hudry, and A. Lobstein. Possible cardinalities for identifying codes in graphs. *Australasian Journal of Combinatorics*, 32 :177–195, 2005.

[38] I. Charon, O. Hudry, and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete mathematics*, 307(3-5) :356–366, 2007.

[39] G. Cohen, I. Honkala, A. Lobstein, and G. Zémor. On identifying codes. In *Proceedings of the DIMACS Workshop on Codes and Association Schemes, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 56, pages 97–109, 2001.

[40] G.D. Cohen, L. Honkala, A. Lobstein, and G. Zémor. On codes identifying vertices in the two-dimensional square lattice with diagonals. *IEEE Transactions on Computers*, 50(2) :174–176, 2001.

[41] C.J. Colbourn, P.J. Slater, and L.K. Stewart. Locating dominating sets in series parallel networks. *Congr. Numer*, 56 :135–162, 1987.

[42] M. Daniel. Codes identifiants. *Mémoire pour le DEA ROCO, Université Joseph Fourier, Grenoble, France*, 2003.

[43] R. Diestel. *Graph theory, volume 173 of Graduate Texts in Mathematics.* Springer, 2005.

[44] A.G. D'yachkov and V.V. Rykov. Bounds on the length of disjunctive codes. *Problemy Peredachi Informatsii*, 18(3) :7–13, 1982.

[45] P. Erdős, J. Pach, R. Pollack, and Z. Tuza. Radius, diameter, and minimum degree. *Journal of Combinatorial Theory Series B*, 47(1) :79, 1989.

[46] P. Erdös, M. Saks, and V.T. Sós. Maximum induced trees in graphs. *Journal of Combinatorial Theory, Series B*, 41(1) :61–79, 1986.

[47] H. Fleischner. The square of every two-connected graph is Hamiltonian. *J. Combinatorial Theory*, 3 :29–34, 1974.

[48] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó, and C. Smyth. Codes identifying sets of vertices in random networks. *Discrete Mathematics*, 307(9-10) :1094–1107, 2007.

[49] MR Garey and DS Johnson. The rectilinear Steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4) :826–834, 1977.

[50] M.R. Garey and D.S. Johnson. *Computers and intractability.* Freeman San Francisco, 1979.

[51] A. Ghouila-Houri. Diamètre maximal d ?un graphe fortement connexe. *Comptes rendus de l'Académie des sciences, Paris*, 250 :254–256, 1960.

[52] S. Gravier, R. Klasing, and J. Moncel. Hardness results and approximation algorithms for identifying codes and locating-dominating codes in graphs. *Algorithmic Operations Research*, 3(1) :43–50, 2008.

[53] S. Gravier and J. Moncel. Construction of codes identifying sets of vertices. *the electronic journal of combinatorics*, 12(R13), 2005.

[54] S. Gravier and J. Moncel. On graphs having a $V \setminus \{x\}$ set as an identifying code. *Discrete mathematics*, 307(3-5) :432–434, 2007.

[55] S. Gravier, J. Moncel, and A. Semri. Identifying codes of cycles. *European Journal of Combinatorics*, 27(5) :767–776, 2006.

[56] TW Haynes, ST Hedetniemi, and PJ Slater. *Fundamental of domination in graphs : Advanced Topics*. Marcel Dekker, Inc, New York, NY, 1998.

[57] TW Haynes, ST Hedetniemi, and PJ Slater. *Fundamentals of domination in graphs*. Marcel Dekker, Inc, New York, NY, 1998.

[58] I. Honkala, M.G. Karpovsky, and S. Litsyn. On the identification of vertices and edges using cycles. *Lecture Notes in Computer Science*, pages 308–314, 2001.

[59] I. Honkala, M.G. Karpovsky, and S. Litsyn. Cycles identifying vertices and edges in binary hypercubes and 2-dimensional tori. *Discrete Applied Mathematics*, 129(2-3) :409–419, 2003.

[60] I. Honkala and T. Laihonen. Codes for identification in the king lattice. *Graphs and Combinatorics*, 19(4) :505–516, 2003.

[61] I. Honkala, T. Laihonen, and S. Ranto. On codes identifying sets of vertices in Hamming spaces. *Designs, Codes and Cryptography*, 24(2) :193–204, 2001.

[62] I. Honkala and A. Lobstein. On identification in $\mathbb{Z}^2$ using translates of given patterns. *Journal of Universal Computer Science*, 9(10) :1204–1219, 2003.

[63] V. Junnila and T. Laihonen. Optimal identifying codes in cycles and paths. *Soumis pour publication.*

[64] R.M. Karp. Reducibility among combinatorial problems. *Complexity of computer computations*, 43 :85–103, 1972.

[65] M.G. Karpovsky, K. Chakrabarty, and L.B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Transactions on Information Theory*, 44(2) :599–611, 1998.

[66] W. Kautz and R. Singleton. Nonrandom binary superimposed codes. *IEEE Transactions on Information Theory*, 10(4) :363–377, 1964.

[67] M. Laifenfeld and A. Trachtenberg. Identifying codes and covering problems. *IEEE Transactions on Information Theory*, 54(9) :3929–3950, 2008.

[68] T. Laihonen. On cages admitting identifying codes. *European Journal of Combinatorics*, 29(3) :737–741, 2008.

[69] T. Laihonen and J. Moncel. On graphs admitting codes identifying sets of vertices. *Australasian Journal of Combinatorics*, 41 :81–91, 2008.

[70] T. Laihonen and S. Ranto. Codes identifying sets of vertices. *Lecture Notes in Computer Science*, pages 82–91, 2001.

[71] L.C. Lau. Bipartite roots of graphs. *ACM Transactions on Algorithms (TALG)*, 2(2) :208, 2006.

[72] Y.L. Lin and S.S. Skiena. Algorithms for square roots of graphs. *SIAM Journal on Discrete Mathematics*, 8(1) :99–118, 1995.

[73] A. Lobstein. Identifying and locating-dominating codes in graphs, a bibliography. *Published electronically at http ://perso. enst. fr/ lobstein/debutBIBidetlocdom. pdf.*

[74] J. Moncel. *Codes identifiants dans les graphes*. PhD thesis, These de Doctorat de l'Université de Grenoble, France, 2005.

[75] J. Moncel. Constructing codes identifying sets of vertices. *Designs, Codes and Cryptography*, 41(1) :23–31, 2006.

[76] J. Moncel. On graphs on n vertices having an identifying code of cardinality $\lceil \log_2(n+1) \rceil$. *Discrete Applied Mathematics*, 154(14) :2032–2039, 2006.

[77] R. Motwani and M. Sudan. Computing roots of graphs is hard. *Discrete Applied Mathematics*, 54(1) :81–88, 1994.

[78] A. Mukhopadhyay. The square root of a graph. *J. Combin. Theory*, 2 :290–295, 1967.

[79] O. Ore. Diameters in graphs. *J. Combin. Theory*, 5 :75–81, 1968.

[80] M. Pelto. New bounds for $(r, \leq 2)$-identifying codes in the infinite king grid. *Cryptography and Communications*, pages 1–7, 2009.

[81] S. Ray, R. Ungrangsi, F. De Pellegrini, A. Trachtenberg, and D. Starobinski. Robust location detection in emergency sensor networks. In *Proceedings of INFOCOM 2003*, volume 2, pages 1044–1053, 2003.

[82] D.L. Roberts and F.S. Roberts. Locating sensors in paths and cycles : the case of 2-identifying codes. *European Journal of Combinatorics*, 29(1) :72–82, 2008.

[83] P. Rosendahl. On the identification of vertices using cycles. *Electronic Journal of Combinatorics*, 10(1), 2003.

[84] P. Rosendahl. On the identification problems in products of cycles. *Discrete Mathematics*, 275(1-3) :277–288, 2004.

[85] J. Ross and F. Harary. The square root of a tree. *Bell System Tech.*, 39 :641–647, 1960.

[86] P.J. Slater. Domination and location in acyclic graphs. *Networks*, 17(1) :55–64, 1987.

[87] P.J. Slater. Dominating and reference sets in a graph. *Journal of Mathematical and Physical Sciences*, 22 :445–455, 1988.

[88] J. Suomela. Approximability of identifying codes and locating–dominating codes. *Information Processing Letters*, 103(1) :28–33, 2007.

[89] M. Xu, K. Thulasiraman, and X.D. Hu. Identifying codes of cycles with odd orders. *European Journal of Combinatorics*, 29(7) :1717–1720, 2008.