



HAL
open science

Densité, VC-dimension et étiquetages de graphes

Sébastien Ratel

► **To cite this version:**

Sébastien Ratel. Densité, VC-dimension et étiquetages de graphes. Mathématique discrète [cs.DM]. Aix-Marseille Université, 2019. Français. NNT : . tel-02436255

HAL Id: tel-02436255

<https://hal.science/tel-02436255>

Submitted on 12 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AIX-MARSEILLE UNIVERSITÉ
ÉCOLE DOCTORALE 184
Laboratoire d'Informatique et Systèmes

Thèse présentée pour obtenir le grade universitaire de
docteur

Discipline : Informatique

Sébastien RATEL

Densité, VC-dimension et étiquetages de graphes

Density, VC-dimension, and graph labelings

Soutenue le 08/11/2019 devant le jury composé de :

Nicolas NISSE	I3S/INRIA	Rapporteur
Laurent VIENNOT	IRIF/INRIA	Rapporteur
Olivier BOUSQUET	Google AI	Examinateur
Nadia CREIGNOU	LIS/AMU	Examinatrice
Cyril GAVOILLE	LaBRI	Examinateur
Nabil MUSTAFA	ESIEE	Examinateur
Victor CHEPOI	LIS/AMU	Directeur de thèse
Arnaud LABOUREL	LIS/AMU	Directeur de thèse

Numéro national de thèse/suffixe local : 2019AIXM0427/024ED184



Cette oeuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](#).

Résumé

Un schéma d'étiquetage est un procédé permettant de distribuer la représentation d'un graphe sur ses sommets. Il consiste en une fonction d'encodage qui attribue des étiquettes binaires (courtes) à chaque sommet, et d'une fonction de décodage qui, étant données les informations locales contenues dans deux étiquettes, et sans connaissance supplémentaire sur le graphe, permet de répondre (rapidement) à un type de requête pré-établi. Une partie des résultats de cette thèse sont initialement motivés par la construction de telles représentations distribuées. Ce document traite cependant de problèmes d'intérêt plus généraux tels que l'étude de bornes sur la densité de graphes, de la VC-dimension de familles d'ensembles, ou de propriétés métriques et structurelles.

Nous établissons dans un premier temps des bornes supérieures sur la densité des sous-graphes de produits cartésien de graphes, puis des sous-graphes de demi-cubes. Pour ce faire, nous définissons des extensions du paramètre classique de VC-dimension (utilisé par Haussler, Littlestone et Warmuth en 1994 pour majorer la densité des sous-graphes d'hypercube). De ces bornes sur la densité, nous déduisons des bornes supérieures sur la longueur des étiquettes attribuées par un schéma d'adjacence à ces deux familles de graphes.

Dans un second temps, nous nous intéressons à des schémas de distance et de routage pour deux familles importantes de la théorie métrique des graphes : les graphes médians et les graphes pontés. Nous montrons que la famille des graphes médians, sans cube, avec n sommets, admet des schémas de distance et de routage utilisant tous deux des étiquettes de $O(\log^3 n)$. Ces étiquettes sont décodées en temps constant pour retourner, respectivement, la distance exacte entre deux sommets, ou le port vers un sommet rapprochant (strictement) une source d'une destination. Nous décrivons ensuite un schéma de distance 4-approché pour la famille des graphes pontés, sans K_4 , avec n sommets, utilisant des étiquettes de $O(\log^3 n)$ bits. Ces dernières peuvent être décodées en temps constant pour obtenir une valeur entre la distance exacte et quatre fois celle-ci.

Mots clés : Arboricité, Demi-cube, Densité, Graphe médian, Graphe ponté, Produit cartésien, Routage, Schéma d'adjacence, Schéma de distance, VC-dimension.

Abstract

A labeling scheme is a way of distributing the representation of a graph over its vertices. Such a scheme consists in an encoding function that constructs a (short) binary label for every vertex, and in a decoding function that can answer (quickly) a predefined query using only the local information contained in two labels (with no further knowledge of the graph). Constructing such distributed representations constituted the initial motivation of most of the results of this document. However, this manuscript concerns problem of more general interest such as bounding the density of graphs, studying the VC-dimension of set families, or investigating on metric and structural properties of graphs.

As a first contribution, we upper bound the density of the subgraphs of Cartesian products of graphs, and of the subgraphs of halved-cubes. To do so, we extend the classical notion of VC-dimension (already used in 1994 by Haussler, Littlestone, and Warmuth to upper bound the density of the subgraphs of hypercubes). From our results, we deduce upper bounds on the size of labels used by an adjacency labeling scheme on these graph classes.

We then investigate on distance and routing labeling schemes for two important families of metric graph theory: median graphs and bridged graphs. We first show that the class of cube-free median graphs on n vertices enjoys distance and routing labeling schemes both using labels of $O(\log^3 n)$ bits. These labels can be decoded in constant time to respectively return the exact distance between two vertices, or a port to take from a source vertex in order to get (strictly) closer to a target one. We then describe an approximate distance labeling scheme for the family of K_4 -free bridged graphs on n vertices. This scheme also uses labels of size $O(\log^3 n)$ that can be decoded in constant time to return a value of at most four times the exact distance between two vertices.

Keywords: Adjacency labeling scheme, Arboricity, Bridged graph, Cartesian product, Density, Distance labeling scheme, Halved-cubes, Median graph, Routing, VC-dimension.

Table des matières

Résumé	3
Abstract	4
Table des matières	5
Introduction générale	9
I Préliminaires et contexte	15
Introduction	16
1 Notions de théorie des graphes	17
1.1 Préliminaires de théorie des graphes	18
1.1.1 Graphes et notations usuels	18
1.1.2 Sous-graphes, distances et intervalles	19
1.1.3 Voisinage, densité et notions connexes	21
1.1.4 Morphismes et plongements de graphes	23
1.2 Produits cartésiens de graphes	24
1.2.1 Hypercube comme produit cartésien	25
1.2.2 Graphes de Hamming	25
1.2.3 Classes de parallélisme et plongement dans l'hypercube	26
1.2.4 Facteurs premiers et irréductibles, plongement canonique	28
1.3 Graphes définis par des familles d'ensembles	30
1.3.1 Hypercube comme graphe d'une famille d'ensembles	30
1.3.2 Demi-cubes et graphes de Johnson	31
2 Dimension de Vapnik-Chervonenkis et densité	33
2.1 VC-dimension et apprentissage	34
2.2 Définitions	36
2.2.1 Lemme de Sauer et classes de concepts maximums	38
2.2.2 Classes de concepts amples et lemme du sandwich	39
2.3 Densité des graphes de 1-inclusion	39
2.4 Preuves classiques du théorème 8	41
2.4.1 Preuve par induction	41

2.4.2	Preuve par décalage	41
2.4.3	Preuve par compression	42
2.5	Généralisations et extensions de VC-dimension	43
3	Schémas d'étiquetage	45
3.1	Introduction	45
3.2	Schémas d'adjacence	46
3.2.1	Arboricité et densité	47
3.2.2	Graphes universels	49
3.3	Schémas de distance	51
3.3.1	Distances dans les arbres	53
3.4	Schémas de routage	54
3.4.1	Routage dans les arbres	55
4	Graphes faiblement modulaires	56
4.1	Graphes faiblement modulaires	56
4.2	Graphes médians	58
4.2.1	Graphes médians et complexes cubiques $CAT(0)$	61
4.2.2	Problèmes de distances dans les graphes médians et les complexes cubiques $CAT(0)$	63
4.3	Graphes pontés	64
4.3.1	Graphes pontés et complexes systoliques	66
4.4	Courbure et théorème de Gauss-Bonnet	67
II	Densité et arboricité	69
	Introduction	70
5	Sous-graphes de produits cartésiens	72
5.1	Résultats principaux	73
5.2	Préliminaires	74
5.2.1	Sous-produits, extensions et fibres	74
5.2.2	Mineurs et sous-produits mineurs	76
5.3	VC-dimension et VC-densité	76
5.4	Preuve du théorème 17	79
5.5	Preuve du théorème 18	80
5.5.1	Propriétés de la VC-dimension et de la VC-densité	81
5.5.2	Preuve du théorème 18	87
5.6	Classes spéciales de graphes	88
5.6.1	Graphes de dégénérescence bornée	89
5.6.2	Graphes de degré moyen poly-logarithmique	89
5.6.3	Produits de graphes démontables	90
5.6.4	Produits de graphes cordaux	91

5.6.5	Produits d'octaèdres	91
5.7	Application aux schémas d'adjacence	93
6	Sous-graphes de demi-cubes	94
6.1	Résultat principal	94
6.2	Préliminaires	95
6.3	Familles d'ensembles plantées et cliques plantées	96
6.4	La clique-VC-dimension	98
6.4.1	La clique-VC-dimension des familles paires plantées	99
6.4.2	La clique-VC-dimension des familles paires quelconques	101
6.5	Preuve du théorème 19	102
6.5.1	d-décalages de familles paires plantées	102
6.5.2	Bouquets de demi-cubes	105
6.5.3	Dégénérescence des bouquets de demi-cubes	106
6.5.4	Preuve du théorème 19	107
6.6	Discussion	108
	Conclusion	111
	III Schémas d'étiquetage de distance et de routage	113
	Introduction	114
7	Graphes médians sans cubes	116
7.1	Résultats principaux	116
7.2	L'idée du schéma	118
7.3	Étoiles dans les graphes médians	119
7.3.1	Schémas de distance et de routage pour les étoiles	120
7.4	Fibres des graphes médians	121
7.5	Fibres dans les graphes médians sans cube	125
7.5.1	Classification des fibres	126
7.5.2	La bordure totale des fibres est quasi-portée	126
7.6	Classification des paires de sommets et plus courts chemins	130
7.7	Schéma de distance pour les graphes médians sans cube	133
7.7.1	Encodage	134
7.7.2	Requêtes de distance	136
7.7.3	Correction et complexité	139
7.8	Schéma de routage pour les graphes médians sans cube	141
7.8.1	L'idée	141
7.8.2	Encodage	142
7.8.3	Requêtes de routage	144
8	Graphes pontés sans K_4	148

8.1	Résultat principal	148
8.2	L'idée du schéma	149
8.3	Propriétés	150
8.4	Étoiles dans les graphes pontés sans K_4	155
8.5	Fibres des graphes pontés sans K_4	156
8.5.1	Frontières et bordures	161
8.5.2	Projections et représentants	162
8.6	Classification des sommets et plus courts chemins	166
8.7	Schéma de distance 4-approché pour les graphes pontés sans K_4	172
8.7.1	Encodage	172
8.7.2	Requêtes de distances	174
8.7.3	Correction et complexité	176
	Conclusion	177
	Conclusion générale	180
	Bibliographie	182
	Index	193
	Table des figures	196

Introduction générale

Les graphes sont un outil mathématique permettant de modéliser un grand nombre de problèmes. Ils consistent en un ensemble d'objets abstraits, que nous appelons des *sommets* (ou des *nœuds*), avec une relation sur ces objets, deux sommets pouvant être reliés ou non par une arête. L'étude de leurs propriétés structurelles ou métriques, de leurs représentations en mémoire, ou d'algorithmes permettant de résoudre les problèmes qu'ils modélisent mènent à autant de questions fondamentales et variées. Une partie des résultats de cette thèse sont initialement motivés par la construction de représentations (distribuées) de graphes. Cela nous a conduit à nous intéresser à des problèmes d'intérêt plus généraux tels que l'étude de bornes sur la densité de graphes, de la VC-dimension de familles d'ensembles, ou de propriétés métriques et structurelles.

Un schéma d'étiquetage est un procédé permettant de répartir la représentation d'un graphe sur ses sommets plutôt que d'avoir une représentation globale à laquelle chaque nœud doit avoir accès. Cette représentation distribuée doit permettre de répondre à une requête fixée à l'avance (adjacence, distance, routage, etc.) en n'utilisant que des informations locales, attribuées aux sommets sous la forme d'étiquettes binaires. Pour qu'une telle représentation soit pertinente, nous souhaitons généralement que les étiquettes attribuées soient courtes (c.-à-d., poly-logarithmique en la taille du graphe) et qu'elles puissent être décodées rapidement (c.-à-d., en temps constant en leur longueur). Sous ces conditions, KANNAN, NAOR et RUDICH [106] ont montré que tous les graphes ne peuvent pas être représentés de façon distribuée de manière efficace. Ainsi, une première étape lorsque nous nous intéressons à des schémas d'étiquetage consiste à trouver une famille suffisamment restreinte pour en faire un "bon candidat" à une représentation distribuée.

En ce qui concerne les schémas d'adjacence, selon une conjecture de KANNAN, NAOR et RUDICH [106], ces bons candidats sont exactement les familles de graphes closes par sous-graphes induits, et possédant moins de $2^{O(n \log n)}$ graphes de n sommets. Pour les schémas de distance ou de routage, il est nécessaire de se concentrer sur des sous-familles encore plus restreintes, avec davantage de propriétés métriques, à moins de considérer des approximations.

Densité et arboricité

Les premiers schémas d'étiquetage considérés portaient sur des requêtes d'adjacence puisque c'est, en quelque sorte, la requête minimale permettant de reconstruire le graphe [38, 106, 120]. Il se trouve que la majorité des schémas d'adjacence utilisent la notion de *densité* (c.-à-d., le ratio maximal du nombre d'arêtes sur le nombre de sommets pris sur tous les sous-graphes d'un graphe) via le concept d'*arboricité* (c.-à-d., nombre de forêts nécessaires pour couvrir toutes les arêtes d'un graphe). En effet, NASH-WILLIAMS [121] a montré qu'un graphe est de faible arboricité si et seulement s'il est de faible densité. De plus, KANNAN, NAOR et RUDICH [106] ont remarqué qu'une famille de graphes (avec n sommets) dont les arêtes peuvent être couvertes par k forêts possède un schéma d'adjacence avec étiquettes de $(k + 1) \lceil \log n \rceil$ bits. Ces deux résultats considérés de concert établissent le lien entre densité et schéma d'adjacence. Ainsi, trouver une borne supérieure sur la densité d'un graphe permet directement de déduire une borne supérieure sur la longueur des étiquettes attribuées par un schéma d'adjacence.

Une famille de graphes peu denses ayant fait l'objet de nombreux travaux est celle des sous-graphes induits d'hypercubes (avec n sommets). Un résultat du folklore majore la densité de ces derniers par $O(\log n)$. En 1994, HAUSSLER, LITTLESTONE et WARMUTH [100] ont fourni une amélioration notoire de ce résultat en utilisant un paramètre nommé VC-dimension (introduit par VAPNIK et CHERVONENKIS [148]). Ils ont montré que la VC-dimension est une borne supérieure pour la densité d'un sous-graphe d'hypercube. Or, la VC-dimension est toujours inférieure (ou égale) à $\log n$, et des exemples dans lesquels l'écart est arbitrairement grand sont faciles à construire.

L'introduction du concept de VC-dimension (et l'utilisation qui en est faite dans [100]) était motivé par la théorie de l'apprentissage et des probabilités, mais son intérêt s'est avéré bien plus général, notamment en géométrie algorithmique [102, 128] ou en théorie des graphes [36, 41]. L'un des premiers résultats utilisant ce paramètre est dû à SAUER [139], donnant une borne sur la taille d'une famille d'ensembles en fonction de sa VC-dimension. Il a donné lieu à l'introduction de familles d'ensembles maximums et, de façon plus indirecte, aux familles d'ensembles amples [33, 21, 113]. En particulier, les familles amples induisent des sous-graphes isométriques d'hypercubes. Elles généralisent aussi les graphes médians qui sont définis par de fortes propriétés métriques, et qui sont centraux en théorie métrique des graphes. Des extensions du paramètre de VC-dimension ou du lemme de Sauer constituent donc des problèmes intéressants, et fréquemment étudiés [1, 41, 101, 123, 133].

Pour notre part, dans la partie II de ce document, nous prouvons deux analogues du théorème de HAUSSLER, LITTLESTONE et WARMUTH [100] sur la den-

sité des sous-graphes d'hypercubes pour deux familles de graphes bien plus générales : les sous-graphes de produits cartésiens de graphes et les sous-graphes de demi-cubes (c.-à-d., des hypercubes dans lesquels les sommets initialement à distance 2 sont reliés). Ces résultats impliquent l'introduction d'extensions de la notion classique de VC-dimension. Le chapitre 5 est basé sur l'article « On density of subgraphs of Cartesian products » [57]. Nous y montrons d'abord le théorème qui suit.

Théorème. *La densité d'un sous-graphe G d'un produit cartésien Γ de graphes connexes est inférieure à $\log |V(G)|$ fois la plus grande densité d'un facteur de Γ .*

Nous prouvons ensuite le résultat plus fort suivant :

Théorème. *La densité d'un sous-graphe G d'un produit cartésien de graphes connexes excluant un mineur fixé H est majorée par la VC-dimension de G multipliée par une constante $\mu(H)$ dépendant de H .*

Nous raffinons ensuite ces résultats pour des familles plus précises (sous-graphes de produits cartésiens de graphes cordaux ou d'octaèdres, entre autres). Nous en déduisons des bornes supérieures sur la longueur nécessaire pour les étiquettes d'un schéma d'adjacence sur ces familles.

Dans le chapitre 6, issu de l'article « On density of subgraphs of halved cubes » [59], nous montrons le résultat suivant :

Théorème. *La densité d'un sous-graphe de demi-cube $G = (V, E)$ est majorée par le carré de sa clique-VC-dimension.*

Schémas de distance et de routage

Des schémas permettant de répondre à d'autres types de requêtes que la simple adjacence ont aussi été considérés, notamment des schémas de distance ou de routage [12, 53, 85, 131]. Mais la question de l'existence d'un adressage compact des sommets d'un graphe dans le but de calculer des distances exactes est un peu plus ancien. GRAHAM et POLLAK [92] posaient ce problème sous la forme de l'existence d'un plongement isométrique d'un graphe quelconque dans un "cube écrasé" ("squashed cube") de la plus petit dimension possible. Un premier résultat impactant sur les schémas de distance est donc la réponse de WINKLER [151] à ce problème puisqu'il montre que *les sommets d'un graphe connexe à n sommets peuvent être étiquetés par des mots de longueur $n - 1$ sur l'alphabet $\{0, 1, *\}$ de sorte que la distance entre deux sommets soit la distance de Hamming entre leurs étiquettes.* Ce résultat montre que tout graphe connexe à n sommets

admet un schéma de distance avec des étiquettes de taille $O(n)$. Comme pour les schémas d'adjacence, pour espérer obtenir des schémas de distance (ou de routage) poly-logarithmiques, il faut considérer des familles de graphes ayant des fortes propriétés métriques.

En fait, pour tous les types de requêtes, les arbres (ou forêts) constituent la classe de graphes la plus étudiée de par leur structure simple et bien connue, et leur prépondérance dans de nombreux domaines ou problèmes. Le premier schéma de distance (ou routage) efficace pour cette famille est dû à PELEG [131]. Il montre que les arbres avec n sommets admettent un schéma de distance utilisant des étiquettes de $O(\log^2 n)$ bits, ce qui est asymptotiquement optimal [13, 85]. Ce résultat est obtenu en remarquant qu'un arbre avec n sommets admet toujours au moins un sommet dont la suppression crée des composantes connexes d'au plus $n/2$ sommets chacune. Plus généralement, un ensemble de sommets dont la suppression crée des composantes connexes de petite taille est un outil classique pour obtenir des schémas de distance (ou routage), appelé *séparateur équilibré*. Par exemple, les graphes planaires sur n sommets admettent un séparateur équilibré de $O(\sqrt{n})$ sommets et un schéma d'adjacence utilisant des étiquettes de $O(\sqrt{n} \log n)$ bits (et une borne inférieure de $\Omega(\sqrt[3]{n})$). Une autre façon d'obtenir des schémas de distance ou de routage utilise la notion de “*hub*” [2, 63] : l'étiquette de chaque sommet u est un ensemble de sommets (les “*hubs*” de u) associés à leur distance à u ; chaque paire de sommet doit alors avoir un “*hub*” en commun.

Il existe aussi des version approchées de schémas d'étiquetage de distance ou de routage dans lesquelles la réponse à la requête est une approximation. Le décodage peut alors fournir une distance avec erreur additive et/ou multiplicative, ou diriger sur un chemin plus ou moins proche du plus court. Un résultat particulièrement important du domaine est dû à THORUP et ZWICK [146]. Il concerne des représentations compactes de graphes quelconques (pondérés) permettant de répondre à des requêtes de distance approchées (erreur multiplicative) en temps constant. THORUP [144] s'intéresse plus particulièrement à ce type de représentation pour les graphes planaires.

Dans la partie III de ce manuscrit, nous nous intéresserons particulièrement à deux familles de graphes définies par de fortes propriétés métriques : les graphes médians et les graphes pontés. Ces deux familles peuvent être vues comme des généralisations des arbres puisque ces derniers sont exactement les graphes médians de dimension 1 ou les graphes pontés sans K_3 . Les graphes médians constituent aussi une sous-classe importante des sous-graphes isométrique d'hypercubes. Et, de même que les graphes pontés, ils admettent un grand nombre de caractérisations dans des domaines divers. Les graphes médians et pontés sont notamment utiles en théorie géométrique des groupes de par leurs équivalences respectives avec les complexes cubiques CAT(0) [49] et les complexes systoliques

[49, 94, 104].

Ces deux familles, bien que très étudiées et possédant de nombreuses caractérisations restent très difficiles à appréhender. En particulier, l'existence de schémas de distance poly-logarithmique pour les graphes médians ou les graphes pontés généraux constituent deux problèmes ouverts intéressants. Dans la partie III, constituée de deux chapitres, nous présentons des schémas de distance et de routage pour la famille des graphes médians sans cube et pour celle des graphes pontés sans K_4 . Ces deux familles, bien que d'apparence restreinte, possèdent déjà une structure assez complexe. Entre autre, elles ne possèdent aucune restriction sur leur mineurs, ce qui en fait des familles de graphes non planaires. Le chapitre 7 a donné lieu à la publication de « Distance labeling schemes for cube-free median graphs » [58]. Nous y concevons un schéma d'étiquetage de distance pour la famille des graphes médians sans cubes. Nous adaptons ensuite ce schéma pour obtenir un schéma de routage. Nous obtenons donc le résultat ci-dessous.

Théorème. *La famille des graphes médians sans cube avec n sommets admet des schémas de distance et de routage (exacts) utilisant tous deux des étiquettes de longueur $O(\log^3 n)$. Ces étiquettes sont attribuées en temps polynomial et décodées en temps constant.*

Pour terminer, dans le chapitre 8 (qui n'a pas encore fait l'objet d'un article), nous décrivons un schéma de distance approché pour la famille des graphes pontés sans K_4 .

Théorème. *La famille des graphes pontés sans K_4 avec n sommets admet un schéma de distance, avec erreur multiplicative d'au plus 4, utilisant des étiquettes de $O(\log^3 n)$ bits. Ces étiquettes sont attribuées en temps polynomial et décodées en temps constant.*

Nos schémas sont basés sur une partition récursive des graphes en ce que nous appelons une étoile et ses fibres. Cette partition n'est pas équivalente à l'outil classique des séparateurs équilibrés, mais possède aussi des parties équilibrées. La technicité des schémas que nous proposons réside dans les propriétés métriques que nous avons dû démontrer, et qui sont nécessaires à leur fonctionnement. De façon informelle, les distances (ou le routage) entre des sommets dans des fibres suffisamment éloignées seront calculées par rapport à l'étoile, cette dernière jouant alors le rôle d'un séparateur. Par contre, pour les fibres proches, nous sommes amenés à étudier précisément leur structure et celle de plus courts chemins entre les sommets.

Structure du document

Ce manuscrit est constitué de trois parties. Dans la première, nous rappelons des définitions de base, des résultats connus, ainsi que le contexte utiles à la compréhension des parties suivantes. Dans la deuxième partie, nous présentons deux analogues du théorème de HAUSSLER, LITTLESTONE et WARMUTH [100] sur la densité des sous-graphes d'hypercubes pour la famille des sous-graphes de produits cartésiens de graphes et pour celle des sous-graphes de demi-cubes. Pour ces deux problèmes de densité, nous étendons la notion classique de VC-dimension. Enfin, dans la dernière partie, nous construisons des schémas de distance et de routage pour la famille des graphes médians sans cube ainsi que pour celle des graphes pontés sans K_4 .

Première partie
Préliminaires et contexte

Introduction

Dans cette première partie, nous introduisons les notions et le contexte nécessaires à la compréhension des contributions que nous présentons dans les deux parties suivantes.

Plus précisément, dans le premier chapitre, nous donnons des définitions de base de théorie des graphes. Nous présentons aussi quelques résultats classiques motivant certaines contributions de la partie II. Notamment le théorème du “plongement isométrique canonique” de GRAHAM et WINKLER [91].

Dans le chapitre 2, nous introduisons la notion de VC-dimension. Nous commençons par la présenter dans le contexte de l’apprentissage qui l’a initialement vue introduite. Nous en donnons ensuite des définitions équivalentes dans le vocabulaire que nous utilisons par la suite. Nous énonçons aussi un résultat classique de HAUSSLER, LITTLESTONE et WARMUTH [100] majorant la densité d’un sous-graphe d’hypercube par sa VC-dimension, puis nous présentons quelques techniques permettant le prouver. Ces schémas de preuves nous seront utiles dans la partie II pour établir des extensions de ce résultat.

Dans le chapitre 3, nous présentons la notion de schéma d’étiquetage qui motive une partie des résultats de ce document. Nous y rappelons des résultats autour des schémas d’étiquetage d’adjacence, de distance et de routage.

Enfin, dans le chapitre 4, nous introduisons la classe des graphes faiblement modulaires. L’importance de cette famille de graphe est expliquée par le fait qu’elle regroupe nombre de classes centrales en théorie métrique des graphes. Parmi celles-ci, nous pouvons citer en particulier les graphes médians et pontés puisqu’ils constitueront les objets d’étude principaux de la partie III. Dans ce quatrième chapitre, nous traitons donc plus spécifiquement de ces deux classes de graphes, et nous donnons quelques unes de leurs nombreuses caractérisations.

Chapitre 1

Notions de théorie des graphes

Sommaire

1.1	Préliminaires de théorie des graphes	18
1.1.1	Graphes et notations usuels	18
1.1.2	Sous-graphes, distances et intervalles	19
1.1.3	Voisinage, densité et notions connexes	21
1.1.4	Morphismes et plongements de graphes	23
1.2	Produits cartésiens de graphes	24
1.2.1	Hypercube comme produit cartésien	25
1.2.2	Graphes de Hamming	25
1.2.3	Classes de parallélisme et plongement dans l'hypercube	26
1.2.4	Facteurs premiers et irréductibles, plongement canonique	28
1.3	Graphes définis par des familles d'ensembles	30
1.3.1	Hypercube comme graphe d'une famille d'ensembles	30
1.3.2	Demi-cubes et graphes de Johnson	31

Ce premier chapitre est consacré à l'introduction de notions classiques et relativement simples de la théorie des graphes (pour de plus amples informations, voir les livres *Graph Theory* [68], *Handbook of Product Graphs* [95] ou *Geometry of Cuts and Metrics* [66] sur lesquels cette introduction est en partie basée). Dans la section 1.1, nous introduisons le vocabulaire et les définitions de base que nous utiliserons par la suite. Dans les deux sections suivantes, nous présentons succinctement deux façons d'obtenir des graphes, témoins de deux sous-domaines de la théorie des graphes. Plus précisément, dans la section 1.2 nous présentons quelques graphes définis par une propriété structurelle : celle donnée par l'opération de produit cartésien de graphes. Dans la section 1.3, nous donnons des familles de graphes définies par des propriétés combinatoires : ce sont des propriétés sur les sommets, traduites par la relation des arêtes. Ces deux aspects de la théorie des graphes seront respectivement reconsidérés dans les chapitres 5 et 6.

1.1 Préliminaires de théorie des graphes

Un *graphe (simple)* G consiste en deux ensembles : un ensemble V de *sommets* et une relation $E \subseteq V \times V$ sur ces sommets, appelée ensemble des *arêtes* de $G = (V, E)$. Lorsque la relation E est *symétrique* (c.-à-d., $(\forall (u, v) \in V^2, (u, v) \in E \Leftrightarrow (v, u) \in E)$), G est dit *non-orienté*. Quand E est *irréflexive* (c.-à-d., $\forall u \in V, (u, u) \notin E$), G est dit *sans boucles*. Un graphe sur un ensemble de sommets fini est dit *fini*, sinon il est dit *infini*.

Remarque 1. Les graphes que nous considérerons dans ce manuscrit seront **simples, non-orientés, sans boucles** et, la plupart du temps, **finis**.

Les arêtes (u, v) d'un graphe $G = (V, E)$ sont généralement notées uv , et u et v sont appelés les *extrémités* de uv . Un graphe avec un unique sommet et aucune arête est dit *trivial*. Lorsqu'une arête uv est dans E , u est dit *adjacent* à v , ou que v est un *voisin* de u . Deux sommets u et v adjacents seront parfois notés $u \sim v$. De même, nous pourrons noter $u \not\sim v$ le fait qu'ils ne soient pas adjacents. Le nombre de voisins d'un sommet u est appelé le *degré* de u et noté $\deg(u)$.

Un *sous-graphe* $H = (V(H), E(H))$ d'un graphe $G = (V(G), E(G))$ est un graphe constitué d'un sous-ensemble de sommets de G et sur lequel est considéré un sous-ensemble des arêtes de G (c.-à-d., $V(H) \subseteq V(G)$ et $E(H) \subseteq E(G) \cap (V(H) \times V(H))$).

1.1.1 Graphes et notations usuels

Nous présentons maintenant quelques graphes particuliers simples et fréquemment considérés. Nous introduisons leurs notations usuelles, nous les réutiliserons souvent par la suite. Ces graphes sont illustrés dans la figure 1.1.

Un *graphe complet* (ou *clique*) de taille k , noté K_k , est un graphe composé de k sommets deux-à-deux adjacents. Un graphe $G = (V, E)$ dont l'ensemble des sommets peut être partitionné en deux parties A et B de sorte que chaque arête de G ait une extrémité dans A et l'autre dans B est dit *biparti*. Plus généralement, G est *k-parti* si V peut être partitionné en k parties V_1, \dots, V_k de sorte que, pour tout $i \in \{1, \dots, k\}$, pour tout $(u, v) \in V_k \times V_k$, $uv \notin E$. Dans ce cas, si toutes les arêtes autorisées par la k -partition $\{V_1, \dots, V_k\}$ sont dans G , ce dernier est dit *k-parti complet* et noté $K_{|V_1|, \dots, |V_k|}$.

Un *chemin de longueur k* est un graphe, noté P_k , avec k sommets v_1, \dots, v_k et dont les arêtes sont $v_1v_2, \dots, v_{k-1}v_k$. Si ce chemin est "fermé" par une arête v_kv_1 , alors le graphe est appelé un *cycle de longueur k* et noté C_k . Si, pour tout couple $(u, v) \in V \times V$ d'un graphe $G = (V, E)$, il existe un chemin (appelé (u, v) -chemin)

reliant u à v dans G , ce dernier est dit *connexe*. Un cycle de longueur k auquel est ajouté un sommet relié à tous les autres est appelé une k -roue, notée W_k . Un graphe n'admettant aucun cycle est appelé une *forêt* (dans le cas général), ou un *arbre* s'il est connexe. Les sommets de degré 1 d'un arbre T sont appelés des *feuilles*. Il est bien souvent utile, dans les arbres, de distinguer un sommet que nous appelons *racine*. Nous regardons alors l'arbre du point de vue de ce sommet distingué r : la *profondeur* $\text{prof}(v)$ d'un sommet $v \in V(T)$ correspond au nombre d'arêtes sur le chemin (unique) entre v et r ; une *branche* de T est un chemin entre la racine et une feuille ; le *parent* de v est l'unique voisin père(v) de v à profondeur $\text{prof}(v) - 1$ (r n'a pas de parent) ; un *fil* u de v est un voisin de v à profondeur $\text{prof}(v) + 1$ (les feuilles n'ont pas de fils) ; un *ancêtre* u de v est un sommet sur une branche contenant v tel que $\text{prof}(u) < \text{prof}(v)$. La *hauteur* d'un arbre est la profondeur maximum sur ses sommets. Remarquons qu'un chemin est un arbre avec au plus deux feuilles. Une *étoile* est un arbre de hauteur inférieure ou égale à 1. Une *chenille* est un chemin avec des nœuds pendants (c.-à-d., des étoiles reliées par un chemin).

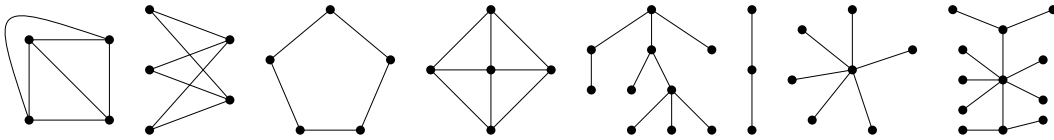


FIGURE 1.1 – De gauche à droite : Un K_4 , un $K_{3,2}$, un C_5 , un W_4 , un arbre (de hauteur 3), un P_3 , une étoile $K_{1,6}$ et une chenille.

1.1.2 Sous-graphes, distances et intervalles

Soit $G = (V, E)$ un graphe (fini, simple, non-orienté et sans boucle).

La *distance* $d_G(u, v)$ entre deux sommets u et v de G correspond au nombre d'arêtes minimal d'un (u, v) -chemin de G . Le *diamètre* $\text{diam}(G)$ de G est la distance maximale entre deux de ses sommets (c.-à-d., $\text{diam}(G) := \max\{d_G(u, v) : (u, v) \in V^2\}$). Pour un sous-graphe H de G , la *projection métrique* $\text{Pr}(u, V(H))$ (ou simplement $\text{Pr}(u, H)$) de u sur $V(H)$ (ou, disons, sur H) correspond à l'ensemble $\{v \in V(H) : \forall v' \in V(H), d_G(u, v') \geq d_G(u, v)\}$ des sommets de H à distance minimum de u . Pour un sous-ensemble $S \subseteq V$ de sommet d'un graphe $G = (V, E)$, la *boule de rayon r autour de S dans G* , notée $B_r(S, G)$, est l'ensemble $\{v \in V : d_G(v, S) \leq r\}$. Lorsque le graphe G est évident d'après le contexte, la boule est simplement notée $B_r(S)$. Aussi, si S est un singleton $\{s\}$, $B_r(\{s\})$ est plutôt notée $B_r(s)$. L'*intervalle* $I(u, v)$ entre u et v consiste en l'ensemble des sommets sur des (u, v) -chemins :

$$I_G(u, v) := \{x \in V : d_G(u, x) + d_G(x, v) = d_G(u, v)\}.$$

En l'absence d'ambiguïté sur le graphe considéré, nous noterons souvent $I(u, v)$ et $d(u, v)$ au lieu de $I_G(u, v)$ et $d_G(u, v)$, respectivement. Une *empreinte* d'un sommet $u \notin V(H)$ sur $V(H)$ est un sommet $v \in V(H)$ tel que $I(u, v) \cap V(H) = \{v\}$. Autrement formulé, une empreinte de $u \notin V(H)$ sur H est un sommet v de H tel qu'aucun plus court chemin entre u et v n'ait de sommet interne dans H . Nous dénotons par $\Upsilon(u, V(H))$ (ou $\Upsilon(u, H)$) l'ensemble des empreintes de u sur $V(H)$. Nous pouvons remarquer que, pour tout sommet u , $\text{Pr}(u, H) \subseteq \Upsilon(u, H)$.

Un graphe H est appelé un sous-graphe *induit* de G (noté $H \subseteq G$), si toutes les arêtes de G ayant leurs deux extrémités dans $V(H)$ sont prises dans $E(H)$ (c.-à-d., $E(H) = E(G) \cap (V(H) \times V(H))$). Pour un sous-ensemble $S \subseteq V(G)$, nous dénoterons parfois par $G(S)$ le sous-graphe de G *induit par* S (c.-à-d., le sous-graphe induit de G dont l'ensemble des sommets est S). La plupart du temps, les sous-graphes que nous considérerons seront (au moins) induits et donc, si le contexte est clair, nous ne ferons bien souvent pas de différence entre un ensemble S et le graphe $G(S)$ qu'il induit.

Un sous-graphe H de G est dit *isométrique* si la distance dans H entre toute paire de sommets de $V(H)$ est la même que celle dans G (c.-à-d., $\forall u, v \in V(H)$, $d_H(u, v) = d_G(u, v)$). Un sous-graphe H de G (ou l'ensemble de sommet $V(H)$) est *convexe* s'il inclue l'intervalle entre toute paire de ses sommets (c.-à-d., $\forall u, v \in V(H)$, $I_G(u, v) \subseteq V(H)$).

Remarque 2. Un sous-graphe convexe de G est isométrique, et un sous-graphe isométrique de G est lui-même induit.

Un sous-graphe H de G (ou l'ensemble de sommet $V(H)$) est *porté* si, pour tout choix d'un sommet $u \in V(G)$, il existe un sommet $u' \in V(H)$ tel que, pour tout $v \in V(H)$,

$$d_G(u, v) = d_G(u, u') + d_G(u', v).$$

Ce sommet u' est alors nommé une *porte* de u sur H . De par sa définition, il est facile de remarquer qu'une porte doit être unique. Les différentes notions de sous-graphes sont illustrées dans la figure 1.2. Remarquons que H est porté si et seulement si l'empreinte de tout sommet $u \notin V(H)$ sur H contient un unique élément. D'après cela, nous dirons qu'un graphe H (ou un ensemble $V(H)$) est *k-porté* si chaque sommet hors de H possède au plus k empreintes sur H (c.-à-d., $\forall u \notin V(H)$, $|\Upsilon(u, V(H))| \leq k$). En particulier, nous dirons que H est *quasi-porté* si $|\Upsilon(u, V(H))| \leq 2$ pour tout $u \notin V(H)$. Précisons que cette notion nous sera essentiellement utile dans la partie III.

Proposition 1. *Un sous-graphe porté H d'un graphe G est nécessairement convexe (mais la réciproque n'est pas toujours vraie).*

Démonstration. Considérons un sous-graphe porté H d'un graphe G ainsi que deux sommets u et v de H . Supposons par l'absurde que H ne soit pas convexe. Alors

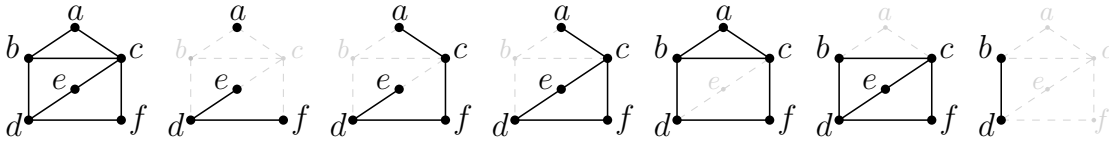


FIGURE 1.2 – De gauche à droite, un graphe G et un sous-graphe quelconque (non connexe), connexe (non induit), induit (non isométrique), isométrique (non convexe), convexe (mais non porté), puis porté de G .

il existe un sommet $x \in I(u, v)$ n'appartenant pas à H . Puisque H est porté, x admet une porte x' sur H , mais alors $d_G(u, x) = d_G(u, x') + d_G(x', x)$ et $d_G(v, x) = d_G(v, x') + d_G(x', x)$. Nous en déduisons que $x = x'$, une contradiction. \square

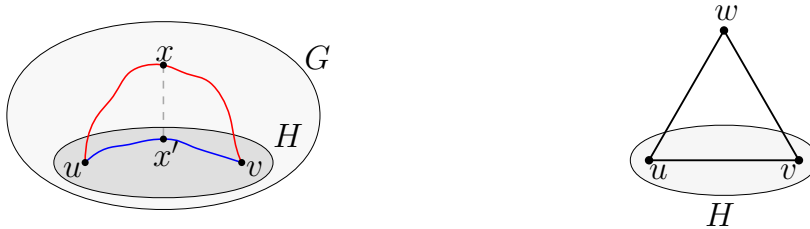


FIGURE 1.3 – Illustration de la proposition 1. Le dessin de gauche est lié à la preuve, celui de droite donne un exemple d'un sous-graphe H convexe mais non porté dans G .

1.1.3 Voisinage, densité et notions connexes

Soit $G = (V, E)$ un graphe. Le *voisinage fermé* dans G d'un sommet $v \in V$, dénoté par $N_G[v]$ (parfois simplement $N[v]$), consiste en v ainsi qu'en tous les sommets qui lui sont adjacents. Le *voisinage ouvert* $N_G(v)$ (ou $N(v)$) de v correspond à $N[v] \setminus \{v\}$. Le nombre $\text{deg}(G) := \frac{1}{|V|} \sum_{v \in G} d(v) = \frac{2|E|}{|V|}$ est appelé *degré moyen* de G . Le *degré moyen maximum* $\text{deg}^*(G)$ de G est le maximum des degrés moyens pris sur tous les sous-graphes G' de G : $\text{deg}^*(G) := \max\{\text{deg}(G') : G' \subseteq G\}$. La *densité* $\text{dens}(G)$ de G est le ratio maximal $|E(G')|/|V(G')|$ pris sur tous les sous-graphes G' de G . Nous pouvons remarquer que les maxima définissant $\text{dens}(G)$ et $\text{deg}^*(G)$ sont en fait atteints pour des sous-graphes induits de G . La densité et le degré moyen maximum sont deux paramètres étroitement liés puisque $\text{dens}(G) = \frac{1}{2} \text{deg}^*(G)$. Nous pouvons faire l'observation simple suivante :

Lemme 1. *Soit G un graphe simple et connexe. Alors G possède au moins deux sommets de degré au plus $\lceil \text{deg}^*(G) \rceil$.*

Démonstration. Supposons, par l'absurde, qu'il existe un graphe connexe G , avec n sommets, admettant un unique sommet v_0 de degré au plus $\beta := \lceil \text{deg}^*(G) \rceil$. Donc

chacun de ses $(n - 1)$ autres sommets a un degré d'au moins $\beta + 1$. Comme G est supposé connexe, $\deg(v_0) \geq 1$. Nous obtenons donc

$$\deg^*(G) \geq \frac{(\beta + 1)(n - 1) + 1}{n} = \beta + 1 - \frac{\beta}{n} > \beta,$$

mais cela contredit $\beta = \lceil \deg^*(G) \rceil$. □

Un graphe G est dit k -dégénéré s'il existe un ordre de démontage v_1, \dots, v_n de ses sommets dans lequel, pour tout $i \in \{1, \dots, n\}$, v_i a un degré d'au plus k dans le sous-graphe $G[v_i, \dots, v_n]$ de G induit par v_i, v_{i+1}, \dots, v_n . Le k minimal tel que G soit k -dégénéré est appelé la *dégénérescence* de G , notée $\delta^*(G)$. Autrement dit,

$$\delta^*(G) := \max_{H \subseteq G} (\min_{v \in V(H)} \deg(v))$$

Vu dans l'autre sens, G est k -dégénéré s'il peut être construit, à partir de l'ensemble vide, par la répétition de l'opération suivante :

Ajouter un sommet et le relier à, au plus, k sommets déjà existants.

Il est bien connu que la dégénérescence d'un graphe $G = (V, E)$ fournit une borne supérieure pour sa densité.

Proposition 2. *Pour tout graphe (simple et sans boucle) G , l'inégalité suivante est vérifiée*

$$\text{dens}(G) := \max \left\{ \frac{|E(H)|}{|V(H)|} : H \text{ sous-graphe induit de } G \right\} \leq \delta^*(G).$$

Démonstration. Soit G un graphe avec n sommets, m arêtes et de dégénérescence k . Alors il existe un ordre de démontage des sommets de G dans lequel nous pouvons toujours retirer un sommet de degré $\leq k$. Pour tout $i \in \{0, n - 1\}$, dénotons par v_i le $(n - i)$ -ième sommet de G retiré dans cet ordre de démontage (c.-à-d., le premier sommet retiré est noté v_{n-1} et le dernier est noté v_0). Nous pouvons remarquer que, pour tout $i \in \{0, \dots, n - 1\}$, le degré de v_i dans $G[v_0, \dots, v_i]$ est d'au plus $\min(i, k)$. Par conséquent,

$$\begin{aligned} m \leq \sum_{i=0}^{n-1} \min(i, k) &= \sum_{i=0}^k i + \sum_{i=k+1}^{n-1} k \\ &= nk - \frac{k(k+1)}{2} \leq nk \end{aligned}$$

La dernière ligne étant due au fait que k soit, par définition, un entier naturel. Il est facile de constater que la dégénérescence d'un graphe G est supérieure ou égale à celle de chacun de ses sous-graphes. Nous en déduisons la propriété annoncée. □

1.1.4 Morphismes et plongements de graphes

Soient $G = (V, E)$ et $G' = (V', E')$ deux graphes. Un *homomorphisme* est une application $f : V \rightarrow V'$ telle que, pour tout $u, v \in V$, si $uv \in E$ alors $f(u)f(v) \in E'$. Un *isomorphisme* est une bijection $g : V \rightarrow V'$ telle que, pour tout $u, v \in V$, $uv \in E$ si et seulement si $g(u)g(v) \in E'$ (c.-à-d., c'est un homomorphisme bijectif dont la réciproque est un homomorphisme). Si une telle bijection existe, G et G' sont dit *isomorphes*. En général, deux graphes isomorphes G et G' sont considérés comme étant les mêmes (puisque alors G' est égal à G à renommage près de ses sommets), nous le notons $G \simeq G'$. Un graphe G est *plongeable isométriquement* dans un graphe G' s'il existe une application $\varphi : V \rightarrow V'$ telle que, pour tout $u, v \in V$, $d_G(u, v) = d_{G'}(\varphi(u), \varphi(v))$. G est *plongeable avec une distorsion* $D \geq 1$ dans G' s'il existe une application $\varphi : V \rightarrow V'$ telle que, pour tout $u, v \in V$,

$$\frac{1}{D} d_{G'}(\varphi(u), \varphi(v)) \leq d_G(u, v) \leq D d_{G'}(\varphi(u), \varphi(v)).$$

Un graphe G est un *rétracté* de G' s'il existe un homomorphisme $\rho : V' \rightarrow V$ et un homomorphisme $\tau : V \rightarrow V'$ tels que $\rho \circ \tau$ corresponde à la fonction identité sur V . L'homomorphisme ρ est appelée une *rétraction* de G' vers G . Autrement dit, un graphe G est un rétracté de G' si $G \subseteq G'$ et s'il existe un homomorphisme de G' vers G envoyant les sommets de G sur eux-mêmes (voir figure 1.4).

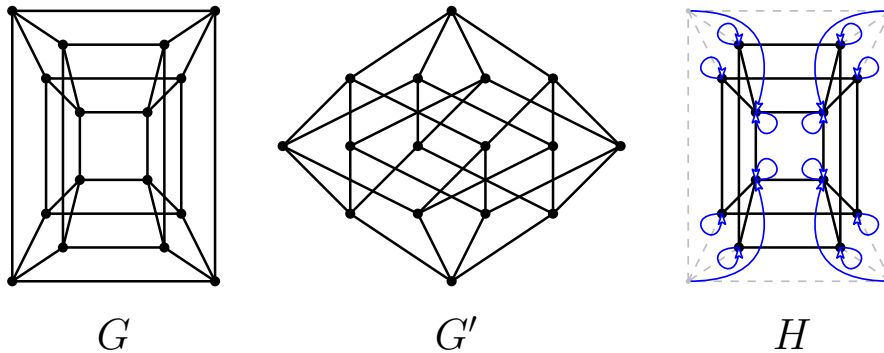


FIGURE 1.4 – Un graphe G , un graphe G' isomorphe à G et un rétracté H de G (les flèches bleues représentent l'opération de rétraction de G vers H , les sommets des H sont en noir). G est isomorphe à un Q_4 (définit en section 1.2.1) et H est un graphe médian de dimension 3 (définit en section 4.2).

1.2 Produits cartésiens de graphes

Le *produit cartésien* $\Gamma := \prod_{i=1}^m G_i =: G_1 \square \cdots \square G_m$ de m graphes G_1, \dots, G_m est le graphe dont les sommets sont l'ensemble des m -uplets $(x_1, \dots, x_m) \in V(G_1) \times \cdots \times V(G_m)$, et dans lequel deux sommets $x = (x_1, \dots, x_m)$ et $y = (y_1, \dots, y_m)$ sont adjacents si et seulement s'il existe un indice $i \in \{1, \dots, m\}$ tel que $x_i y_i \in E(G_i)$ et $x_j = y_j$ pour tout $j \neq i$. Pour un facteur G_i fixé, et pour une arête uv de G_i , toutes les arêtes de Γ reliant deux sommets de formes respectives $(x_1, \dots, x_{i-1}, u, x_{i+1}, \dots, x_m)$ et $(x_1, \dots, x_{i-1}, v, x_{i+1}, \dots, x_m)$ sont appelées *arêtes de type uv* (voir la figure 1.5). De même, les extrémités de ces arêtes sont respectivement appelées des *sommets de type u* ou de type v .

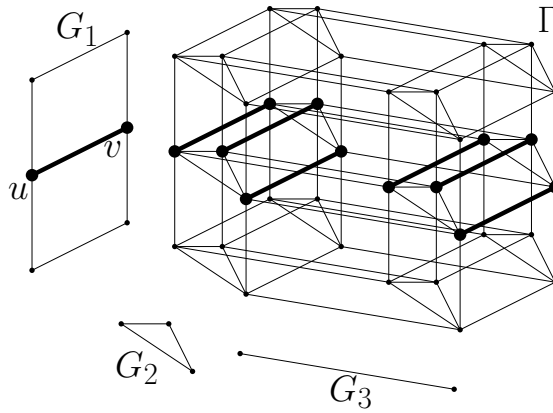


FIGURE 1.5 – Exemple d'un produit cartésien $\Gamma := G_1 \square G_2 \square G_3$ et d'une arête $uv \in E(G_1)$ avec ses copies dans Γ (c.-à-d., les arêtes de type uv de Γ).

Le résultat suivant est un simple exercice de dénombrement, mais il sera utile par la suite.

Proposition 3. Soit $\Gamma := G_1 \square \cdots \square G_m$ un produit cartésien de m graphes $G_1 = (V_1, E_1), \dots, G_m = (V_m, E_m)$. Alors Γ possède :

- $|V_1| \cdot \dots \cdot |V_m|$ sommets ;
- $\sum_{i=1}^m \left(|E_i| \cdot \prod_{\substack{1 \leq j \leq m \\ j \neq i}} |V_j| \right)$ arêtes.

Démonstration. Les sommets de Γ sont tous les vecteurs de $V_1 \times \cdots \times V_m$, il y en a bien $|V_1| \cdot \dots \cdot |V_m|$. Le nombre d'arêtes de Γ correspond au nombre de façons de fixer toutes les coordonnées sauf la i -ième ($i \in \{1, \dots, m\}$), et de choisir une arête $uv \in E_i$. Pour un $i \in \{1, \dots, m\}$ donné, il y a $\prod_{1 \leq j \leq m, j \neq i} |V_j|$ manières de fixer chaque coordonnée sauf la i -ième, et $|E_i|$ arêtes dans G_i . D'où le résultat. \square

1.2.1 Hypercube comme produit cartésien

L'hypercube Q_m de dimension m (ou m -cube), illustré en figure 1.4, peut être défini de nombreuses façons. Nous le définissons ici comme le produit cartésien de m copies du graphe complet K_2 (c.-à-d., $Q_m = K_2 \square \cdots \square K_2$). La dimension m de Q_m pourra aussi être dénotée $\dim(Q_m)$. Nous donnerons quelques définitions équivalentes dans des sections suivantes. Le résultat ci-dessous est bien connu (voir, par exemple, [95, Lemme 3.2] ou [90]) :

Théorème 1. Soit $G = (V, E)$ un sous-graphe induit d'hypercube. Alors

$$\frac{|E|}{|V|} \leq \log_2 |V|. \dagger$$

Démonstration. Soit m la dimension de l'hypercube Q dont G est un sous-graphe. Montrons l'inégalité par induction sur le nombre de sommets de G , le cas de $|V| = 1$ est trivial. Supposons donc $|V| \geq 2$ et choisissons n'importe quel facteur K_2 de Q et nommons u et v ses sommets. Nous pouvons constater que les arêtes de type uv induisent une coupe de G (c.-à-d., leur suppression déconnecte G). Posons A le côté de la coupe formé par les sommets de type u , B celui formé par ceux de type v , et $E(A, B)$ l'ensemble des arêtes de la coupe. Alors $|V| = |V(A)| + |V(B)|$ et l'un de A ou de B doit contenir au plus la moitié des sommets de G . Supposons $|V(A)| \leq |V|/2$. Nous considérons les hypothèses d'inductions suivantes : $|E(A)| \leq |V(A)| \log |V(A)|$ et $|E(B)| \leq |V(B)| \log |V(B)|$. Nous avons

$$\begin{aligned} |E| &= |E(A)| && + && |E(B)| && + && |E(A, B)| \\ &\leq |V(A)| \log |V(A)| && + && |V(B)| \log |V(B)| && + && |V(A)| \\ &\leq |V(A)| \log(|V|/2) && + && |V(B)| \log |V| && + && |V(A)| \\ &= |V| \log |V| \end{aligned}$$

□

Notons que cette borne est atteinte si G est un hypercube, mais qu'elle peut être arbitrairement éloignée de la vérité.

1.2.2 Graphes de Hamming

Un *graphe de Hamming* est un produit cartésien de cliques de tailles quelconques. Un graphe de Hamming produit de d cliques de taille k est noté $H(d, k)$. L'hypercube de dimension d est donc le cas particulier du graphe de Hamming $H(d, k)$.

†. Dans tout ce manuscrit, les logarithmes de base non spécifiée seront supposés en base 2.

Théorème 2. Soit $G = (V, E)$ un sous-graphe induit du graphe de Hamming $\Gamma := K_{k_1} \square \cdots \square K_{k_m}$ et soit $\omega := \max\{k_i : 1 \leq i \leq m\}$. Alors

$$\frac{|E|}{|V|} \leq \omega \log_{\omega} |V|.$$

Démonstration. La preuve est tout à fait similaire à celle du théorème 1. Nous choisissons ici n'importe quel facteur K_{k_j} de Γ ainsi qu'un sommet u de ce facteur. Nous posons alors A l'ensemble des sommets de type u de G , B l'ensemble des sommets restant, et $E(A, B)$ l'ensemble des arêtes entre A et B . Nous pouvons remarquer que A ne peut pas contenir plus de $|V|/\omega$ sommets (car il correspond au sous-graphe de G dans une des ω copies de $K_{k_1} \square \cdots \square K_{k_{j-1}} \square K_{k_{j+1}} \square \cdots \square K_{k_m}$). De plus, chaque sommet de A peut être voisin d'au plus ω sommets de B . Les mêmes calculs que précédemment conduisent au résultat annoncé. \square

Remarque 3. Dans le chapitre 5 (section 5.4), nous montrerons une généralisation des théorèmes 1 et 2 aux sous-graphes de produits cartésiens de graphes quelconques.

1.2.3 Classes de parallélisme et plongement dans l'hypercube

La *relation de parallélisme* Θ , a été introduite par DJOKOVIĆ [69] pour le cas des graphes bipartis, puis elle a été étendue aux graphes quelconques par WINKLER [152]. Cette relation joue un rôle important dans la caractérisation structurelle des sous-graphes isométriques ou convexes d'hypercubes, ainsi que dans les plongements de graphes dans des produits cartésiens (comme nous le verrons dans la section suivante). Commençons par voir la notion introduite par DJOKOVIĆ pour démontrer le théorème suivant :

Théorème 3 (DJOKOVIĆ [69]). *Un graphe $G = (V, E)$ est un sous-graphe isométrique d'hypercube si et seulement si G est biparti, connexe et $W_G(u, v) := \{x \in V : d_G(x, u) < d_G(x, v)\}$ est convexe pour tout choix de $uv \in E$.*

Considérons un graphe biparti $G = (V, E)$. Deux arêtes uv et xy sont en *relation* Θ si et seulement si xy a une extrémité dans $W_G(u, v)$ et l'autre dans $W_G(v, u)$.

Remarque 4. Lorsque G est un sous-graphe isométrique d'hypercube, Θ est la clôture transitive de la relation Θ' telle que $uv\Theta'xy$ si $uvxy$ est un C_4 (carré) de G .

Lorsque G est un sous-graphe d'hypercube Q , les classes d'équivalence par Θ définissent des coupes. En fait, chacune de ces coupes détermine un facteur dans l'hypercube canonique. Si, de plus, G est isométrique dans Q , alors pour toute arête uv , la coupe induite par la classe d'équivalence $\Theta(uv)$ contenant uv

sépare G en deux ensembles convexes $\{W_G(u, v), W_G(v, u)\}$. Ces deux ensembles complémentaires $W_G(u, v)$ et $W_G(v, u)$ sont alors appelés des *demi-espaces*. Réciproquement, pour chaque séparation de G en deux ensembles convexes, il existe une arête uv telle que $\{W_G(u, v), W_G(v, u)\}$ corresponde à cette séparation. Nous dénotons par $\{\Theta_i : i \in I = \{1, \dots, \dim(Q)\}\}$ l'ensemble des classes d'équivalences de la relation Θ , appelées *classes de parallélisme* par BARTHÉLEMY et CONSTANTIN [26] (voir la figure 1.6, droite). Pour une classe d'équivalence Θ_i , nous dénotons par $\{A_i, B_i\}$ la séparation convexe associée. Lorsque $u \in A_i$ et $v \in B_i$, nous disons que Θ_i sépare les sommets u et v . Le plongement isométrique φ de G dans un hypercube est alors obtenu en choisissant un sommet de base x de G qui sera envoyé sur \emptyset par φ ($\varphi(x) := \emptyset$) puis, pour chaque autre sommet y , $\varphi(y)$ consiste en l'ensemble des classes de parallélisme de Θ qui séparent y de x . C'est-à-dire :

$$\begin{aligned} \varphi : V &\rightarrow 2^I \\ y &\mapsto \{i : \Theta_i \text{ sépare } x \text{ de } y\} \end{aligned}$$

Nous pouvons remarquer que ce plongement dans l'hypercube peut en fait être effectué pour tout graphe biparti dans lequel, pour chaque arête uv , $W_G(u, v)$ et $W_G(v, u)$ sont convexes. Et c'est exactement ainsi que DJOKOVIĆ [69] caractérise les graphes admettant un plongement isométrique dans l'hypercube dans le théorème 3 (voir figure 1.6, droite).

Par la suite, la relation Θ a été étendue par WINKLER [152] des graphes bipartis aux graphes quelconques de la façon suivante : $uv\Theta xy$ si et seulement si $d_G(u, x) + d_G(v, y) \neq d_G(u, y) + d_G(v, x)$ (où xy et uv sont au sein d'une même composante connexe). Θ^* correspond alors à la clôture transitive de Θ . Dans le cas des sous-graphes d'hypercubes, les relations Θ de WINKLER et de DJOKOVIĆ coïncident et, de plus, $\Theta = \Theta^*$. Cependant, la relation de WINKLER est plus générale. En effet, $uv\Theta^* xy$ si et seulement si xy possède une extrémité dans $W_G^-(u, v) := \{x \in V : d_G(x, u) = d_G(x, v)\}$ et l'autre dans $W_G(u, v)$ ou dans $W_G(v, u)$ (voir l'illustration 1.6, gauche).

Remarque 5. Lorsque G est un sous-graphe isométrique de graphe de Hamming, Θ correspond alors à la fermeture transitive de Θ'' qui met en relation les arêtes uv et xy étant soit opposées dans un carré, soit dans un même triangle.

WINKLER [152] donne alors un analogue pour les produits cartésiens de K_3 au théorème 3 pour les produits cartésiens de K_2 :

Théorème 4 (WINKLER [152]). *Un graphe $G = (V, E)$ est un sous-graphe isométrique de $H(d, 3)$ (pour un certain $d \in \mathbb{N}^*$) si et seulement si $\Theta = \Theta^*$.*

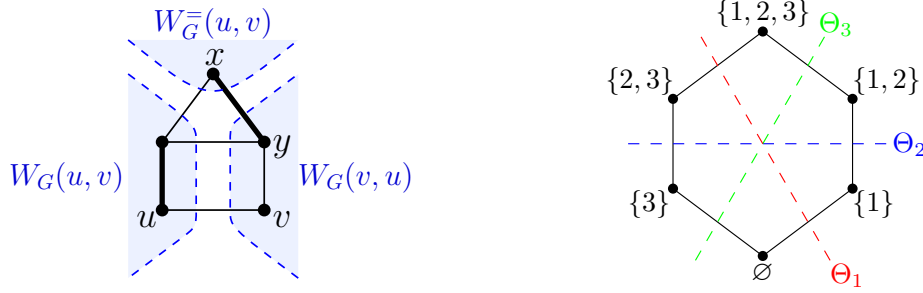


FIGURE 1.6 – À gauche, deux arêtes (uv et xy) en relation Θ . Ici $W_G(u, v) = W_G^-(x, y)$, $W_G(v, u) = W_G(y, x)$ et $W_G^-(u, v) = W_G(x, y)$. À droite, un plongement isométrique d'un C_6 dans un Q_3 .

CHEPOI [46] a ensuite caractérisé entièrement les sous-graphes isométriques de graphes de Hamming.

Théorème 5 (CHEPOI [46]). *Un graphe $G = (V, E)$ est un sous-graphe isométrique de graphe de Hamming si et seulement si, pour toute arête $uv \in E$, les ensembles $W_G(u, v)$, $W_G(v, u)$ et $W_G^-(u, v)$ ainsi que leurs complémentaires sont convexes.*

1.2.4 Facteurs premiers et irréductibles, plongement canonique

Un graphe G est dit *premier* s'il est non-trivial et qu'il ne peut pas être représenté comme un produit cartésien de deux graphes non-triviaux. Autrement dit, pour toute représentation de G comme $G_1 \square G_2$, G_1 ou G_2 doit être un K_1 .

Proposition 4. *Tout graphe non-trivial G admet une décomposition en un produit cartésien d'au plus $\log_2 |V(G)|$ graphes premiers.*

Démonstration. Un graphe non-trivial contient au moins deux sommets donc, si nous posons k le nombre de facteurs non-triviaux dans la décomposition de G , nous obtenons $2^k \leq |V(G)|$. \square

Si G est un graphe quelconque, sa décomposition en produit cartésien de graphes premiers n'est pas toujours unique (voir [95, chapitre 6] et figure 1.7). Elle l'est lorsque G est connexe selon le résultat suivant montré indépendamment par SABIDUSSI [138] et VIZING [149].

Théorème 6 (SABIDUSSI-VIZING [138, 149]). *Tout graphe connexe et fini G admet une unique représentation en produit cartésien de facteurs premiers (à isomorphisme et ordre des facteurs près).*

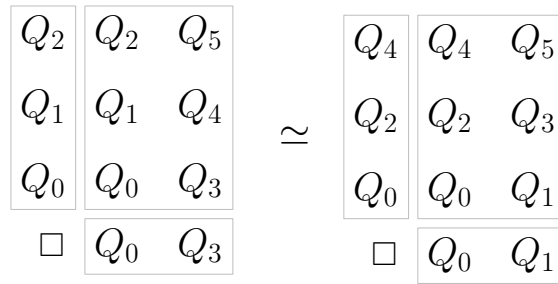


FIGURE 1.7 – Factorisation (non unique) en produit de facteurs premiers. Exemple de deux graphes (non connexes) isomorphes (composés de Q_i déconnectés pour $i = 0, \dots, 5$), admettant deux factorisations distinctes en produit de facteurs premiers : $(Q_0 \cup Q_1 \cup Q_2) \square (Q_0 \cup Q_3)$ et $(Q_0 \cup Q_2 \cup Q_4) \square (Q_0 \cup Q_1)$.

Un plongement isométrique $\varphi : G \rightarrow G_1 \square \dots \square G_k$ est *irrédundant* si, pour tout $i \in \{1, \dots, k\}$, G_i n'est pas trivial et si, pour tout $v \in V(G_i)$, G contient un sommet de type v dans son plongement. Autrement dit, les facteurs sont non-triviaux et chacun de leur sommet est nécessaire (sur au moins une copie) pour représenter G . Un graphe G est *irréductible* si, pour tout plongement irrédundant de G dans un produit cartésien $G_1 \square \dots \square G_k$, l'un des G_i est isomorphe à G et tous les autres sont triviaux. Remarquons qu'un graphe irréductible est nécessairement premier, mais que la réciproque n'est pas vraie.

Théorème 7 (GRAHAM et WINKLER [91], plongement isométrique canonique). *Pour tout graphe fini G , il existe un ensemble G_1, \dots, G_k de facteurs irréductibles et un plongement isométrique de G dans $G_1 \square \dots \square G_k$, et ces facteurs sont uniques.*

La preuve du théorème 7 (détaillée dans les livres de DEZA et LAURENT [66] ou HAMMACK, IMRICH et KLAUŽAR [95]) donne aussi un algorithme efficace pour calculer le plongement canonique. En effet, notons $\Theta_1^*, \dots, \Theta_k^*$ les Θ^* -classes d'arêtes d'un graphe fini $G = (V, E)$. Pour chaque Θ^* -classe Θ_i^* , nous posons $G_i := (V, E \setminus \Theta_i^*)$. Pour une partition \mathcal{P} des sommets de G , le *graphe quotient* G/\mathcal{P} est le graphe ayant \mathcal{P} comme ensemble de sommets et dans lequel deux sommets P et P' sont reliés si et seulement s'il existe une arête de G entre P et P' . Pour chaque G_i , nous posons $\mathcal{P}_i := \{P_1^i, \dots, P_{t_i}^i\}$ sa partition en composantes connexes maximales et nous définissons l'application $\varphi_i : V \rightarrow \mathcal{P}_i$ par $\varphi_i(v) := P_j^i$, avec $v \in P_j^i$. Alors l'application

$$\begin{aligned} \varphi : V &\rightarrow \mathcal{P}_1 \times \dots \times \mathcal{P}_k \\ v &\mapsto (\varphi_1(v), \dots, \varphi_k(v)) \end{aligned}$$

Définit un plongement isométrique canonique de G dans le produit cartésien de graphes irréductibles $G/\mathcal{P}_1 \square \dots \square G/\mathcal{P}_k$ (voir la figure 1.8).

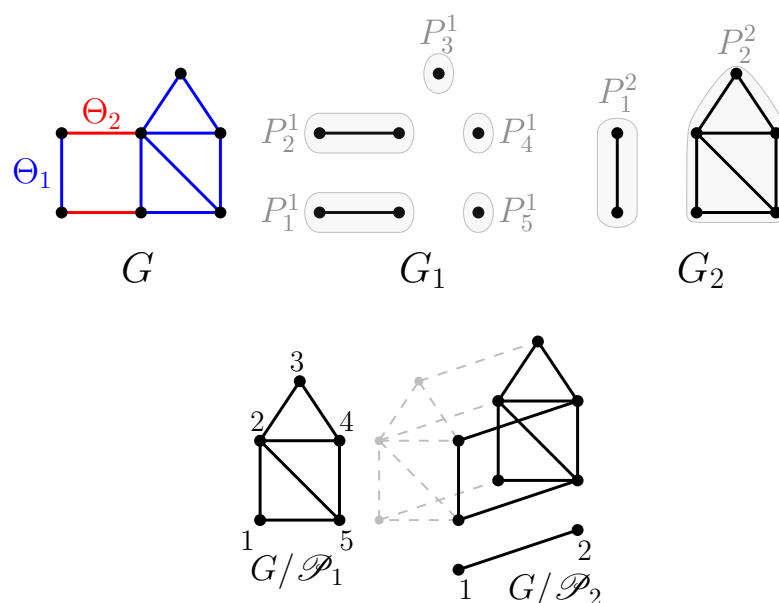


FIGURE 1.8 – Exemple du plongement isométrique canonique d'un graphe G .

Du théorème 7, nous obtenons directement le corollaire suivant :

Corollaire 1. *Un graphe peut être plongé isométriquement dans un graphe de Hamming si et seulement si chaque facteur de sa représentation métrique canonique (celle donnée par le théorème 7) est un graphe complet.*

1.3 Graphes définis par des familles d'ensembles

Dans cette section, X dénotera un univers avec m éléments. Nous présentons ici quelques familles de graphes dont les sommets sont des sous-ensembles de X et dont l'ensemble des arêtes est induit par une relation combinatoire entre ces sous-ensembles. Précisons avant de commencer que, dans un souci de concision, un ensemble de cardinal k sera souvent nommé un k -ensemble. Rappelons aussi que la *différence symétrique* $S \Delta S'$ de deux ensembles S et S' est l'ensemble $(S \cup S') \setminus (S \cap S')$.

1.3.1 Hypercube comme graphe d'une famille d'ensembles

En plus de la définition en termes de produits cartésiens, nous pouvons définir l'*hypercube* Q_m de dimension m à l'aide d'ensembles. C'est le graphe dont l'ensemble des sommets est l'ensemble des parties de X et dans lequel deux sommets S et S' sont reliés par une arête si et seulement si $|S \Delta S'| = 1$. Nous

pouvons alors voir toute famille $\mathcal{S} \subseteq 2^X$ comme un sous-ensemble des sommets de Q_m . Réciproquement, à tout sous-graphe G de Q_m nous pouvons associer une famille d'ensembles $\mathcal{S} \subseteq 2^X$ telle que G soit le sous-graphe d'hypercube dont \mathcal{S} est l'ensemble des sommets. Nous pourrions alors dénoter G par $G(\mathcal{S})$ (ou encore par $G_1(\mathcal{S})$, en référence au graphe de 1-inclusion dont la définition suit). Les sous-graphes d'hypercubes sont aussi nommés “*cubical graph*” par GAREY et GRAHAM [78], ou encore *graphe de 1-inclusion* (de \mathcal{S}) par HAUSSLER [99] et HAUSSLER, LITTLESTONE et WARMUTH [100] dans le cadre de la théorie de l'apprentissage. Les graphes de 1-inclusion, ont de nombreuses applications dans ce domaine, notamment en stratégies de prédictions ou en conception de schémas de compressions, nous aborderons cela dans le chapitre 2.

Une famille \mathcal{S} de sous-ensembles de X (de *simplexes*) est un *complexe simplicial (abstrait)* si elle contient l'ensemble vide \emptyset et qu'elle est close par inclusion (c.-à-d., $\forall S \in 2^X, \forall S' \subseteq S, S \in \mathcal{S} \Rightarrow S' \in \mathcal{S}$). Les ensembles S de cardinal k de \mathcal{S} sont appelés des *k-simplexes*. Et l'ensemble des sous-ensembles S' d'un simplexe S est appelé une *face*. Un *bouquet de cubes* (que nous appellerons aussi *étoile* dans le chapitre 7), est un sous-graphe d'hypercube induit par un complexe simplicial.

1.3.2 Demi-cubes et graphes de Johnson

Le *demi-cube* $\frac{1}{2}Q_m$ est le graphe dont les sommets sont les sous-ensembles de cardinal pair de X et ayant pour arêtes les paires de sommets AB telles que $|A\Delta B| = 2$ (évidemment, considérer les sous-ensembles impairs de X donnerait le même résultat). Notons que $\frac{1}{2}Q_m$ est aussi isomorphe au *carré* Q_{m-1}^2 de l'hypercube Q_{m-1} (c.-à-d., le graphe dont les sommets sont ceux de Q_{m-1} et où deux sommets sont reliés si et seulement s'ils sont à distance 1 ou 2 dans Q_{m-1}). Pour un entier $r > 0$, le *graphe de Johnson* $J(r, m)$ possède les sous-ensembles de cardinal r de X comme sommets et deux sommets A et B sont reliés dans $J(r, m)$ si et seulement si $|A\Delta B| = 2$. Remarquons que, pour tout $r \leq m$, $J(r, m)$ est un sous-graphe isométrique de $\frac{1}{2}Q_m$. Remarquons aussi que le demi-cube $\frac{1}{2}Q_m$ (et donc les graphes de Johnson $J(r, m)$) sont plongeables dans Q_m avec une échelle facteur 2.

Les classes des demi-cubes et des graphes de Johnson contiennent de nombreuses sous-classes importantes en théorie métrique des graphes (voir, par exemple, « *Metric graph theory and geometry : a survey* » [23]). En particulier, les graphes de bases de matroïdes sont des sous-graphes isométriques de graphes de Johnson (MAURER [116]) et les graphes de bases des delta-matroïdes pairs sont des sous-graphes isométriques de demi-cubes (CHEPOI [50]). Des classes plus générales encore que celles-ci sont celles des graphes admettant un plongement isométrique dans les graphes de Johnson ou les demi-cubes. Nous pouvons

noter ici que, de même que les sous-graphes isométriques d'hypercube ont été caractérisés par DJOKOVIĆ [69], CHEPOI [51] a caractérisé les sous-graphes isométriques de graphes de Johnson (en utilisant une relation similaire à Θ et le concept d'“*atom graph*” introduit par SHPECTOROV [141]). Quant au problème de donner une caractérisation similaire des sous-graphes isométriques de demi-cubes, il a aussi été posé dans *Geometry of Cuts and Metrics* [66], mais il reste ouvert. SHPECTOROV [141] a démontré que les graphes admettant un plongement isométrique dans les espaces ℓ_1 sont exactement les graphes admettant un plongement avec échelle constante dans un hypercube. Il a de plus montré que ces graphes sont exactement les sous-graphes isométriques de produits cartésiens dont les facteurs sont des octaèdres (que lui-même nomme des “*cocktail party graphs*”) et des sous-graphes isométriques de demi-cubes. La plupart de ces résultats, sont rappelés dans le livre de DEZA et LAURENT [66].

Remarquons que, de même que nous avons définis des bouquets de cubes, nous pouvons définir un *bouquet de demi-cubes* comme l'ensemble des sous-ensembles de cardinal pair d'un complexe simplicial. Cette définition sera rappelée et utilisée dans le chapitre 6 (section 6.5.2).

Chapitre 2

Dimension de Vapnik-Chervonenkis et densité

Sommaire

2.1	VC-dimension et apprentissage	34
2.2	Définitions	36
2.2.1	Lemme de Sauer et classes de concepts maximums	38
2.2.2	Classes de concepts amples et lemme du sandwich	39
2.3	Densité des graphes de 1-inclusion	39
2.4	Preuves classiques du théorème 8	41
2.4.1	Preuve par induction	41
2.4.2	Preuve par décalage	41
2.4.3	Preuve par compression	42
2.5	Généralisations et extensions de VC-dimension	43

Le concept de VC-dimension a été introduit par VAPNIK et CHERVONENKIS [148] dans le cadre de la théorie des probabilités. La notion s’est avérée d’un intérêt direct en théorie de l’apprentissage (classification binaire, dans un premier temps) et s’est par la suite révélée utile dans des domaines autour de la combinatoire. Dans la section 2.1, nous introduisons la notion de VC-dimension dans le contexte de l’apprentissage qui l’a initialement motivée. Cette première section se veut en partie informelle car elle traite de sujets, important pour la mise en contexte, mais qui ne sont pas au cœur de cette thèse. La section 2.2 suivante est consacrée à l’introduction de la VC-dimension dans des vocabulaires plus proches de nos propres motivations. Nous présentons aussi des familles d’ensembles particulièrement étudiées (les classes de concepts maximums et amples), ainsi que les résultats fondamentaux (lemmes de Sauer et du “sandwich”) à l’origine de leurs définitions. Dans la section 2.3, nous présentons des résultats faisant le lien entre la densité des sous-graphes d’hypercubes (bornée par leur VC-dimension, théorème 8) et des stratégies de prédiction ou des schémas de compression. Dans la section suivante 2.4, nous rappelons trois preuves classiques du théorème 8.

Nous serons amenés à utiliser des techniques similaires dans les chapitres 5 et 6. Enfin, nous terminons en évoquant quelques travaux portant sur des généralisations de la notion de VC-dimension, dans la section 2.5.

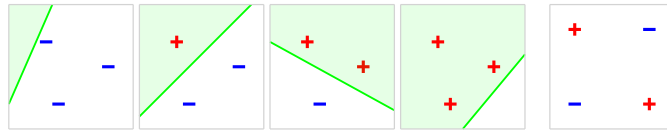
2.1 VC-dimension et apprentissage

Un problème d'*apprentissage* (de *classification*) consiste à trouver une fonction f , dans une classe de fonctions \mathcal{F} définies sur un *domaine* X et à valeur dans un *codomaine* (discret) Y . Précisons qu'en règle générale, X et Y ne sont pas nécessairement supposés finis, mais pour ce qui nous intéresse dans cette introduction et pour la suite, nous considérerons qu'ils le sont. En fait, nous nous intéresserons essentiellement au problème de la *classification binaire*, dans lequel $Y = \{0, 1\}$.

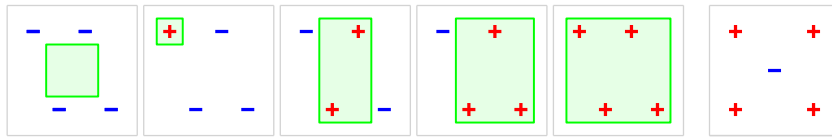
Dans ce contexte, nous avons donc $X := \{x_1, \dots, x_n\}$ ($n \in \mathbb{N}$), et la description d'une fonction f quelconque peut être vue comme un vecteur c de $\{0, 1\}^n$ tel que $c_i = 0$ si et seulement si $f(x_i) = 0$. En apprentissage combinatoire, de tels vecteurs sont appelés des *concepts*. Nous pouvons être amenés à faire un certain nombre d'hypothèses sur l'espace de fonctions \mathcal{F} dans lequel f est recherchée. Ces hypothèses faites sur \mathcal{F} peuvent alors être représentées par un ensemble de parties $\mathcal{C} \subseteq \{0, 1\}^n$ (\mathcal{C} peut aussi être vu comme un ensemble de fonctions indicatrices). \mathcal{C} est alors appelé une *classe de concepts* et peut être vue comme un sous-graphe d'hypercube, alors appelé graphe de 1-inclusion de \mathcal{C} (c.-à-d., les éléments de $\{0, 1\}^n$ peuvent être considérés comme les vecteurs caractéristiques des ensembles de 2^X , c'est d'ailleurs plutôt en tant que parties de X que nous verrons nos classes de concepts).

Intuitivement, la VC-dimension exprime la complexité des hypothèses faites pour rechercher notre fonction de prédiction, une VC-dimension faible est caractéristique d'hypothèses fortes (voir l'exemple 1). En termes un peu plus précis (les détails sont donnés ensuite), un cube Q est dit *pulvérisé* par une classe de concepts \mathcal{C} si, sur la restriction de ses concepts aux coordonnées qui définissent Q , la classe de concepts permet toutes les prédictions (dichotomies) possibles. La VC-dimension correspond à la plus grande dimension d'un cube pulvérisé par notre classe de concepts. Plus précisément, pour une classe de concepts $\mathcal{C} \subseteq 2^X$ et un sous-ensemble $S \subseteq X$, la *trace* de \mathcal{C} sur S (ou *restriction* de \mathcal{C} à S) correspond à l'ensemble $\mathcal{C}|_S := \{C \cap S : C \in \mathcal{C}\}$. S est dit *pulvérisé* par \mathcal{C} si la trace de \mathcal{C} sur S vaut l'ensemble 2^S de ses parties (c.-à-d., $\forall S' \subseteq S, \exists C \in \mathcal{C}$ t.q. $C \cap S = S'$). La *dimension de Vapnik-Chervonenkis* [148, 76] $\text{VC-dim}(\mathcal{C})$ de \mathcal{C} correspond au cardinal du plus grand sous-ensemble de X pulvérisé par \mathcal{C} (c.-à-d., $\text{VC-dim}(\mathcal{C}) := \max\{|S| : S \subseteq X \text{ et } \mathcal{C}|_S = 2^S\}$).

Exemple 1. Un exemple très classique consiste à considérer X comme un ensemble de points du plan euclidien, à séparer en deux classes $Y = \{+, -\}$ en n'utilisant que des droites. C'est-à-dire que \mathcal{F} est l'ensemble des demi-plans de \mathbb{R}^2 , et un élément de X sera étiqueté "+" par $f \in \mathcal{F}$ s'il appartient au demi-plan définissant f , "-" sinon. Alors, l'ensemble de trois points donné sur la figure ci-dessous (gauche) est pulvérisé. En revanche, aucun ensemble de quatre points ne peut être pulvérisé par \mathcal{F} , ce qui montre que sa VC-dimension vaut 3 (un ensemble de quatre points étiquetés non séparables par une droite est donné à droite).



Maintenant, si \mathcal{F} correspond à l'ensemble des rectangles axes-parallèles, alors la figure ci-dessous (gauche) montre un ensemble de quatre points pulvérisés. Cependant, cinq points ne peuvent pas être pulvérisés, la VC-dimension de \mathcal{F} vaut donc 4 (un ensemble de cinq points étiquetés non séparables par des rectangles axes-parallèles est donné à droite de la figure).



Exemple 2. Donnons maintenant un exemple dans lequel la partie "apprentissage" a été abstraite. Le domaine X dénote un univers fini $\{1, 2, 3, 4, 5\}$, notre classe de concepts \mathcal{C} est l'ensemble des vecteurs caractéristiques des ensembles $\{\emptyset, \{2, 3\}, \{1, 3, 4\}, \{1, 4, 5\}, \{1, 2, 3, 4\}\}$. Alors l'ensemble $\{1, 2\}$ est pulvérisé par \mathcal{C} . En effet, la restriction des ensembles de \mathcal{C} aux éléments de $\{1, 2\}$ donne l'ensemble des parties de $\{1, 2\}$: $\{\emptyset, \{2, 3\}, \{1, 3, 4\}, \{1, 4, 5\}, \{1, 2, 3, 4\}\}$.

Pour sélectionner une fonction f (dans un modèle de classification \mathcal{F}) prédisant la catégorie d'objets "inconnus", un algorithme d'apprentissage \mathcal{A} doit disposer d'une base d'objets "connus". C'est-à-dire d'un échantillon d'apprentissage (ou ensemble d'objets étiquetés) $A \subseteq X \times Y$. Cet échantillon permet à l'algorithme \mathcal{A} de calculer une estimation (empirique) de l'erreur (réelle) de classification faite par f sur les éléments de X . Nous ne rentrerons pas dans les détails (notamment du calcul de l'erreur, de ce qu'est un algorithme d'apprentissage ou de comment est choisi un échantillon d'apprentissage) car cela nous éloigne du sujet principal de ce document. Pour des précisions, le lecteur pourra consulter « Théorie de l'apprentissage » [35] ou *Machine learning : a theoretical approach* [122].

Notons seulement que la VC-dimension permet, entre autre de borner la taille de l'échantillon d'apprentissage nécessaire pour avoir un taux d'erreur "satisfaisant" en fonction des hypothèses dans \mathcal{C} . Cette borne peut être obtenue en utilisant un outil de géométrie algorithmique : les ε -réseaux¹. Pour un concept *cible* $C \in \mathcal{C}$ duquel l'algorithme d'apprentissage \mathcal{A} cherche à se rapprocher, l'erreur faite par un concept *hypothèse* $H \in \mathcal{C}$ (produit par \mathcal{A}) est donnée par $|H \Delta C|$. Les *régions d'erreurs* par rapport à C sont définies par l'ensemble $\mathcal{C} \Delta C := \{H \Delta C : H \in \mathcal{C}\}$, dont nous pouvons montrer que $\text{VC-dim}(\mathcal{C} \Delta C) = \text{VC-dim}(\mathcal{C})$ (nous le verrons plus tard, dans un autre contexte, en proposition 10). Pour un $\varepsilon > 0$ donné, $\mathcal{C} \Delta_\varepsilon C$ dénote l'ensemble $\{H \Delta C : H \in \mathcal{C}, |C \Delta H| > \varepsilon\}$ des régions de "forte erreur". Un sous-ensemble $S \subseteq X$ est un ε -réseau pour $\mathcal{C} \Delta C$ s'il touche chacune de ses régions de forte erreur (c.-à-d., $\forall R \in \mathcal{C} \Delta_\varepsilon C, S \cap R \neq \emptyset$). En d'autres termes, si les exemples de l'échantillon d'apprentissage forment un ε -réseau, l'erreur faite par le concept choisi par \mathcal{A} sera faible. Un résultat important de VAPNIK et CHERVONENKIS [148], amélioré à plusieurs reprise, notamment par HAUSSLER et WELZL [102] établit que toute classe de concepts de VC-dimension d doit avoir un ε -réseau de taille $O(\frac{d}{\varepsilon} \log \frac{d}{\varepsilon})$. Ce résultat peut donc être utilisé pour calculer la probabilité que les exemples d'apprentissage forment un ε -réseau. En conséquence, il est possible de borner supérieurement le nombre d'exemples nécessaires à l'apprentissage d'un concept faisant peu d'erreur.

2.2 Définitions

Les notations que nous considérons à partir d'ici changent un peu de celles utilisées dans la section précédente, car nous définissons maintenant la VC-dimension dans le cadre de la combinatoire et la théorie des graphes. À partir d'ici, $X = \{x_1, \dots, x_n\}$ est un univers dont n dénote le nombre d'éléments, $\mathcal{S} \subseteq 2^X$ est une famille de parties de X , et Y dénotera généralement un sous-ensemble de X .

Comme nous l'avons vu brièvement dans la section précédente, la VC-dimension est définie via une opération de pulvérisation. Nous donnons maintenant, plusieurs façons de voir (définir) cette opération. Une *extension* dans 2^X d'un sous-ensemble Y_0 de Y est un ensemble de la forme $S = Y_0 \cup Z$, avec $Z \in X \setminus Y$. La *fibre* $F_Y(Y_0)$ (ou simplement $F(Y_0)$) de Y_0 dans 2^X correspond à l'ensemble de ses extensions. C'est-à-dire, $F(Y_0) := \{Y_0 \cup Z : Z \in X \setminus Y\}$. La figure 2.1 donne une schématisation des fibres et extensions. Un sous-ensemble $Y \subseteq X$ est *pulvérisé* par \mathcal{S} si \mathcal{S} contient une extension de chaque $Y_0 \subseteq Y$ ou, autrement

1. Pour des applications de la VC-dimension et des ε -réseaux en géométrie algorithmique, le lecteur pourra consulter le livre de PACH et AGARWAL [128].

dit, si chaque sous-ensemble de Y possède une fibre non-vide dans \mathcal{S} . Alors

$$\text{VC-dim}(\mathcal{S}) := \max\{|Y| : Y \subseteq X \text{ et } \forall Y_0 \subseteq Y, F_Y(Y_0) \cap \mathcal{S} \neq \emptyset\}.$$

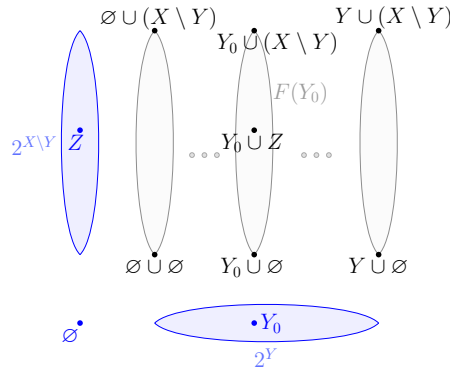


FIGURE 2.1 – Exemple schématique de fibres et d’extensions, vues dans une sorte de repère centré en \emptyset , dont l’abscisse et l’ordonnée sont respectivement indicées par les parties de Y et de $X \setminus Y$.

Si l’on considère \mathcal{S} comme un sous-ensemble de sommets de l’hypercube $Q_n := K_2 \square \cdots \square K_2$ (produit cartésien de n arêtes), l’opération de pulvérisation peut être redéfinie en termes plus proches de la théorie des graphes de la façon qui suit. Pour un ensemble $Y \subseteq X := \{1, \dots, n\}$, posons $Q|Y$ le produit cartésien des facteurs de $Q_n = K_2 \square \cdots \square K_2$ indexés par Y ($Q|Y$ est un $|Y|$ -cube). Alors un ensemble $Y \subseteq X$ est pulvérisé par \mathcal{S} (ou par $G(\mathcal{S})$, le sous-graphe d’hypercube induit par \mathcal{S}) si la contraction des facteurs indexés par $X \setminus Y$ dans $G(\mathcal{S})$ produit $Q|Y$. Ici, la contraction d’un facteur $K_2 := \{u, v\}$ dans $G(\mathcal{S})$ revient à contracter chaque arête de type uv dans Q_n en un nouveau sommet w , et à conserver w si et seulement si l’une (au moins) des extrémités de l’arête l’ayant produit était dans $G(\mathcal{S})$. Une illustration est donnée en figure 2.2.

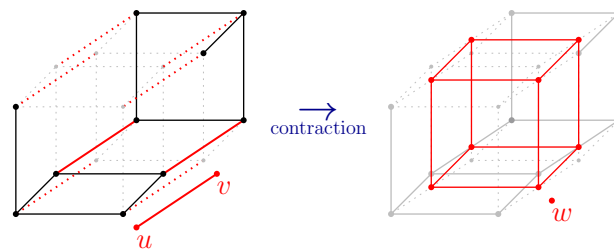


FIGURE 2.2 – Exemple d’un sous-graphe G (en noir) de l’hypercube $Q_4 := K_2 \square K_2 \square K_2 \square K_2$ (dont les arêtes manquantes dans G sont en pointillés). La contraction des arêtes de type uv (en rouge) produit un Q_3 . La VC-dimension de G (ou de $V(G)$) est donc égale à 3.

Lorsque nous calculons la VC-dimension d’une famille d’ensembles vue comme

sous-graphe d'hypercube, il est nécessaire de supposer que ce plongement dans l'hypercube est donné, car ce dernier a un impact sur la VC-dimension. Par exemple, si G est le chemin P_5 , la VC-dimension de G pourra différer selon que G soit plongé dans Q_4 ou dans Q_3 . Plus précisément, si P_5 est plongé dans un 4-cube tel qu'il commence en \emptyset et termine en $\{1, 2, 3, 4\}$, alors $\text{VC-dim}(P_5) = 1$. En revanche, si P_5 est plongé dans un 3-cube de telle sorte qu'il commence en \emptyset et termine en $\{1, 2\}$, alors $\text{VC-dim}(P_5) = 2$. Cet exemple est illustré en figure 2.3. Cette supposition d'avoir accès au plongement est aussi essentielle d'un point de vue calculatoire dans la mesure où déterminer si un graphe est plongeable dans l'hypercube est un problème NP-complet d'après un résultat de AFRATI, PAPADIMITRIOU et PAPAGEORGIOU [5].

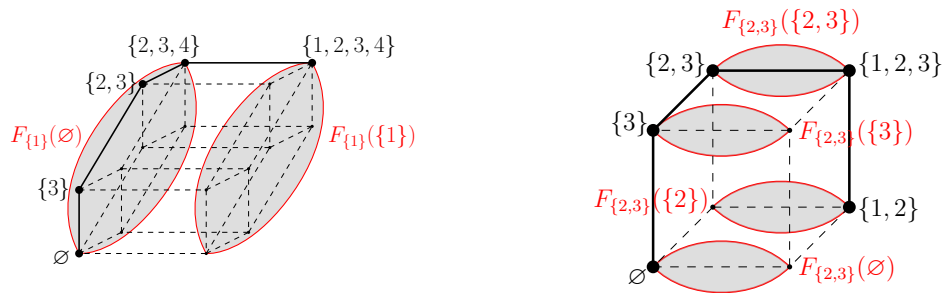


FIGURE 2.3 – Impact du plongement sur la VC-dimension. À gauche, un plongement de P_5 dans Q_4 de VC-dimension 1 ($\{1\}$ est pulvérisé). A droite, un plongement de P_5 dans Q_3 de VC-dimension 2 (l'ensemble $\{2, 3\}$ est pulvérisé).

2.2.1 Lemme de Sauer et classes de concepts maximums

Le lemme fondamental de SAUER [139] (démontré aussi indépendamment par SHELAH [140] et VAPNIK et CHERVONENKIS [148]) est l'un des premiers résultats liés à la VC-dimension. Il donne une borne supérieure serrée sur le cardinal d'une famille d'ensembles \mathcal{S} de VC-dimension fixée d .

Lemme 2 (SAUER [139]). *Pour tout univers X à $n \in \mathbb{N}$ éléments, et pour toute famille $\mathcal{S} \subseteq 2^X$,*

$$|\mathcal{S}| \leq \sum_{i=0}^{\text{VC-dim}(\mathcal{S})} \binom{n}{i}.$$

Une preuve similaire à celle du théorème 8 que nous présentons dans la section 2.4.1 permet aussi de démontrer (par induction) le lemme de Sauer 2.

Une famille $\mathcal{S} \subseteq 2^X$ telle que pour tout $Y \subseteq X$, $\text{VC-dim}(\mathcal{S} \cup \{Y\}) > \text{VC-dim}(\mathcal{S})$ est appelée *maximale*. Si, de plus, le cardinal de \mathcal{S} est égal à la borne supérieure fournie par le lemme de Sauer, alors \mathcal{S} est dite *maximum*.

2.2.2 Classes de concepts amples et lemme du sandwich

Pour un univers $X = \{x_1, \dots, x_n\}$ et une famille \mathcal{S} de sous-ensembles de X , dénotons par $\overline{\mathbf{X}}(\mathcal{S}) := \{Y \subseteq X : \mathcal{S}|_Y = 2^Y\}$ l'ensemble de sous-ensembles de X pulvérisés par \mathcal{S} . De même, posons $\underline{\mathbf{X}}(\mathcal{S}) := \{Y \subseteq X : F(Y) \cap \mathcal{S} = F(Y)\}$ l'ensemble des ensembles de \mathcal{S} induisant des hypercubes dans $G(\mathcal{S})$. Alors le lemme du sandwich suivant est vérifié.

Lemme 3 (Lemme du sandwich). $|\underline{\mathbf{X}}(\mathcal{S})| \leq |\mathcal{S}| \leq |\overline{\mathbf{X}}(\mathcal{S})|$.

Une classe de concept \mathcal{S} est *ample* (ou *extrémale*, ou “*lopsided*”) si $|\mathcal{S}| = |\overline{\mathbf{X}}(\mathcal{S})|$ (c.-à-d., \mathcal{S} pulvérise autant de cubes qu'elle possède d'éléments). Ces classes amples ont été indépendamment découvertes à plusieurs reprises en géométrie, en théorie de l'apprentissage ou en combinatoire. LAWRENCE [113] est le premier à les avoir introduites, sous le nom de “*lopsided sets*”, en étudiant la structure de l'ensemble $\mathcal{S} \subseteq \{-, +\}^d$ des vecteurs de signes d'ensembles convexes de \mathbb{R}^d intersectant certains (des 2^d) orthants en évitant les autres. Cela n'est possible que si \mathcal{S} est “*lopsided*”. Par la suite, BOLLOBÁS, LEADER et RADCLIFFE [33] et BOLLOBÁS et RADCLIFFE [34] introduisent le terme de *classes extrémales* et s'intéressent au lemme de Sauer “*inversé*” : la partie gauche du lemme du sandwich. Ils montrent en particulier l'équivalence entre $|\underline{\mathbf{X}}(\mathcal{S})| = |\mathcal{S}|$ et $|\mathcal{S}| = |\overline{\mathbf{X}}(\mathcal{S})|$ (c.-à-d., intuitivement, chaque cube pulvérisé par \mathcal{S} l'est par un cube induit de $G(\mathcal{S})$). DRESS [71] et BANDELT et al. [21], quant à eux, nomment ces classes *amples* et donnent une importante liste de leurs propriétés. Ils remarquent aussi l'équivalence avec les ensembles “*lopsided*” de LAWRENCE [113].

L'importance des familles amples en combinatoire s'explique aussi par le fait qu'elles généralisent en particulier trois structures centrales : les familles maximums, les antimatroïdes et les graphes médians (ces derniers sont définis dans le chapitre 4).

2.3 Densité des graphes de 1-inclusion

En 1994, HAUSSLER, LITTLESTONE et WARMUTH [100] proposent une stratégie de prédiction pour les classes de concepts basée sur leur graphe de 1-inclusion qu'ils appellent “*1-inclusion prediction strategy*” et qu'ils présentent comme une approche naturelle pour déterminer des modèles de prédiction. Ils obtiennent une borne supérieure pour le risque (dans le pire des cas) de cette stratégie pour une classe de concepts \mathcal{S} dépendant de la densité (dégénérescence) du graphe $G(\mathcal{S})$. Ils établissent en fait une borne serrée pour la densité de $G(\mathcal{S})$ en utilisant la dimension de Vapnik-Chervonenkis $\text{VC-dim}(\mathcal{S})$ de \mathcal{S} .

C'est le théorème 8 suivant (dont plusieurs démonstrations seront données dans la section 2.4).

Théorème 8. *Soit $\mathcal{S} \subseteq 2^X$ une famille de sous-ensembles d'un univers (fini) X et soit $G(\mathcal{S}) = (V, E)$ son graphe de 1-inclusion. Alors*

$$\frac{|E|}{|V|} \leq \text{VC-dim}(\mathcal{S}).$$

Remarquons que l'inégalité $\text{VC-dim}(\mathcal{S}) \leq \log |\mathcal{S}|$ est toujours vérifiée. En effet, pour pulvériser un sous-ensemble de cardinal d , \mathcal{S} doit contenir au moins 2^d éléments. Il s'en suit que $|\mathcal{S}| \geq 2^{\text{VC-dim}(\mathcal{S})}$, et donc que $\log |\mathcal{S}| \geq \text{VC-dim}(\mathcal{S})$. Or, pour certaines classes de concepts, $\text{VC-dim}(\mathcal{S})$ est bien inférieure à $\log |\mathcal{S}|$ et donc $\frac{|E|}{|V|} \leq \text{VC-dim}(\mathcal{S})$ constitue une amélioration notoire du résultat classique $\frac{|E|}{|V|} \leq \log |V|$.

De ce résultat, HAUSSLER, LITTLESTONE et WARMUTH [100] déduisent que le graphe de 1-inclusion peut être dirigé de telle sorte que le degré sortant de tout sommet soit inférieur ou égal à la VC-dimension. LITTLESTONE et WARMUTH [114] font de plus la conjecture qu'une classe de concepts \mathcal{S} de VC-dimension d admet un schéma de compression de taille d . C'est-à-dire, qu'il suffit d'un certain échantillon de taille d (c.-à-d., d exemples étiquetés) pour qu'un algorithme d'apprentissage puisse trouver une fonction prédisant correctement toutes étiquettes. Cela nécessite que le graphe de 1-inclusion associé soit toujours de dégénérescence $\leq d$. Malheureusement, RUBINSTEIN, BARTLETT et RUBINSTEIN exhibent dans « Shifting : One-inclusion mistake bounds and sample compression » [137] un sous-graphe d'hypercube dont le degré minimal est supérieur à la VC-dimension, invalidant la conjecture. Récemment, PÁLVÖLGYI et TARDOS [129] ont aussi donné un exemple de classe de concepts \mathcal{S} n'admettant pas de schéma de compression de longueur $\text{VC-dim}(\mathcal{S})$. Cependant, les contre-exemples donnés ne sont ni des classes maximums, ni des classes amples, ce qui maintient la conjecture pour ces familles. KUZMIN et WARMUTH [112] publient des résultats validant la conjecture pour les classes maximums. La construction qu'ils utilisent pour cela est cependant invalidée par CHALOPIN et al. [44], mais ces auteurs proposent une construction alternative (valide) dans le même article. Ceci-dit, KUZMIN et WARMUTH [112] établissent un lien intéressant entre schéma de compression et démontage par des coins (c.-à-d., démontage par des sommets n'appartenant qu'à un unique cube) d'une classe maximum \mathcal{S} . Ils montrent que l'existence d'un tel démontage doit impliquer un schéma de compression de longueur $\text{VC-dim}(\mathcal{S})$. La question de savoir si les familles amples admettent un tel schéma de compression est, en revanche, laissée ouverte par CHALOPIN et al. [44].

2.4 Preuves classiques du théorème 8

Dans cette section, nous rappelons trois techniques de preuve permettant d'obtenir le théorème 8. Nous commençons par rappeler la preuve par induction proposée en 1994 par HAUSSLER, LITTLESTONE et WARMUTH [100]. Nous continuons avec la preuve alternative, par décalage, fournie un an plus tard par HAUSSLER [99] qui cherchait alors à majorer le nombre de boules disjointes de rayon de r/m pouvant être placées dans un sous-graphe G de Q_m (en centrant ces dernières sur les sommets de G). Son résultat pouvant alors être vu comme une généralisation du lemme de Sauer. Enfin, nous donnons les idées de la preuve de HARPER [97] (1964) et BEZRUKOV [28] (1999), basée sur une opération de compression, pour caractériser les solutions des inégalités arêtes-isopérimétriques dans les hypercubes.

2.4.1 Preuve par induction

Pour un univers fini X et $e \in X$, posons $\mathcal{S}_e := \{S' \subseteq X \setminus \{e\} : \exists S \in \mathcal{S} \text{ t.q. } S' = S \cap X\}$ et $\mathcal{S}^e := \{S' \subseteq X \setminus \{e\} : S', S' \cup \{e\} \in \mathcal{S}\}$. Les preuves du théorème 8, fournie par HAUSSLER, LITTLESTONE et WARMUTH [100], et du lemme de Sauer 2 peuvent alors toutes deux se faire par induction, en se basant sur le lemme fondamental suivant concernant une famille $\mathcal{S} \subseteq 2^X$ de VC-dimension d .

Lemme 4. $|\mathcal{S}| = |\mathcal{S}_e| + |\mathcal{S}^e|$; $\text{VC-dim}(\mathcal{S}_e) \leq d$; et $\text{VC-dim}(\mathcal{S}^e) \leq d - 1$.

Pour $e \in X$, dénotons par G_e et G^e les sous-graphes de l'hypercube $G(2^{X \setminus \{e\}})$ induits respectivement par \mathcal{S}_e et \mathcal{S}^e . Alors $|E(G_e)| \leq d|V(G_e)| = d|\mathcal{S}_e|$ et $E(G^e) \leq (d - 1)|V(G^e)| = (d - 1)|\mathcal{S}^e|$ d'après le lemme 4 et les hypothèses d'induction. Le graphe G_e est obtenu depuis G en contractant les arêtes de l'ensemble F des arêtes de type e dans G . L'ensemble des sommets de G^e est en bijection avec F et deux sommets de G^e sont adjacents si et seulement si l'arête correspondante de F appartient à un carré contenant ces deux sommets. Par le lemme 4, $|V(G)| = |\mathcal{S}| = |\mathcal{S}_e| + |\mathcal{S}^e| = |V(G_e)| + |V(G^e)|$. Les arêtes de G qui, par leur contraction, mènent à des boucles ou des arêtes multiples, et qui sont donc supprimées dans G_e , sont récupérées dans G^e ou dans F . Par conséquent, $|E(G)| \leq |E(G_e)| + |E(G^e)| + |F| = |E(G_e)| + |E(G^e)| + |V(G^e)|$. En nous basant sur cette inégalité, et après quelques calculs, nous obtenons que $\frac{|E(G)|}{|V(G)|} \leq d$.

2.4.2 Preuve par décalage

Pour une famille d'ensembles $\mathcal{S} \subseteq 2^X$, l'opération de *décalage* (“*shifting*”, “*push-down*” ou encore “*stabilization*” dans la littérature) φ_e définie relativement

à un élément $e \in X$ remplace chaque ensemble S de \mathcal{S} tel que $S \setminus \{e\} \notin \mathcal{S}$ par $S \setminus \{e\}$. Nous dénotons par $\varphi_e(\mathcal{S})$ la famille d'ensemble résultante de l'application de φ_e à chaque ensemble de \mathcal{S} . Nous dénotons aussi par $G' := G_1(\varphi_e(\mathcal{S})) = (V', E')$ le graphe de 1-inclusion de $\varphi_e(\mathcal{S})$. HAUSSLER [99] démontre que l'opération de décalage φ_e satisfait les propriétés suivantes :

- (1) φ_e est bijective sur les ensembles de sommets : $|V| = |V'|$,
- (2) φ_e ne diminue pas le nombre d'arêtes : $|E| \leq |E'|$,
- (3) φ_e n'augmente pas la VC-dimension : $\text{VC-dim}(\mathcal{S}) \geq \text{VC-dim}(\varphi_e(\mathcal{S}))$.

Toute famille d'ensembles $\mathcal{S} \subseteq 2^X$ peut être transformée, par une séquence finie de décalages, en une famille \mathcal{S}^* telle que $\varphi_e(\mathcal{S}^*) = \mathcal{S}^*$ pour n'importe quel choix de $e \in X$. La famille \mathcal{S}^* est dite obtenue par *une séquence complète de décalages*, et nous pouvons montrer que c'est un complexe simplicial (et donc, que son graphe de 1-inclusion est un bouquet de cubes).

Posons maintenant $G^* := G_1(\mathcal{S}^*) = (V^*, E^*)$ et $d^* := \text{VC-dim}(\mathcal{S}^*)$. Puisque tous les décalages satisfont les conditions (1), (2) et (3), nous concluons que $|V^*| = |V|$, $|E^*| \geq |E|$ et $d^* \leq d := \text{VC-dim}(\mathcal{S})$. Ainsi, pour démontrer l'inégalité $\frac{|E|}{|V|} \leq d$, il suffit de prouver que $\frac{|E^*|}{|V^*|} \leq d^*$. Pour montrer cela, HAUSSLER utilise le lemme de SAUER [139]. Ceci-dit, cette inégalité peut être obtenue plus directement en bornant la dégénérescence du bouquet de cubes G^* . En effet, dénotons par v_0 le sommet de G^* correspondant à l'origine \emptyset , et considérons un sommet v de G^* parmi les plus éloignés de v_0 . Alors v_0 et v définissent un cube de dimension maximale (et majorée par d^*) dans G^* et, de plus, v n'appartient qu'à ce cube maximal. De ce fait, supprimer v de G^* supprime au plus d^* arêtes et crée un nouveau bouquet de cubes $G^+ = (V^+, E^+)$ avec un sommet de moins et une dimension d'au plus d^* . Par conséquent, nous pouvons appliquer à G^+ nos hypothèses d'induction pour déduire que $|E^+| \leq |V^+|d^*$. Nous obtenons alors que $|E^*| \leq d^* + |E^+| \leq d^* + (|V^*| - 1)d^* = |V^*|d^*$.

2.4.3 Preuve par compression

Le *problème arête-isopérimétrique* pour les hypercubes [28, 97] consiste à trouver, pour tout entier $1 \leq n \leq 2^m$, un sous-graphe G à n sommets d'un m -cube Q_m avec un nombre minimal d'arêtes traversant sa frontière ∂G (c.-à-d., un minimum d'arêtes avec une extrémité dans G et une dans son complément dans Q_m). Les hypercubes étant des graphes réguliers, minimiser ∂G revient à maximiser le nombre d'arêtes de G , et donc sa densité $\frac{|E(G)|}{|V(G)|}$. Un résultat classique de HARPER [97] caractérise les solutions du problème arête-isopérimétrique pour les hypercubes : *pour tout n , c'est le sous-graphe de Q_m induit par les premiers mots binaires de taille n selon l'ordre lexicographique*. HARPER [98] définit une opération de "compression" envoyant les sommets de G sur ces premiers mots binaires

(voir *Global Methods for Combinatorial Isoperimetric Problems* [98, section 3.3]). Il s'avère que les sous-graphes ainsi générés sont en fait des résultats possibles d'une séquence complète de décalages. Plus précisément, HARPER [98, Section 2.4.3] définit une opération, très proche de celle de décalage, qu'il nomme opération de Steiner. Les *opérations de Steiner* sont les applications $\varphi : 2^X \rightarrow 2^X$ satisfaisant les conditions (1) et (2) de la section 2.4.2, ainsi que la condition supplémentaire

$$(4) \ S \subseteq T \text{ implique } \varphi(S) \subseteq \varphi(T).$$

HARPER [97] montre alors que son opération de compression est une opérations de Steiner. Cela permet d'obtenir le théorème 8.

Remarque 6. Une opération de décalage φ_e satisfait aussi (4) mais n'est définie que sur \mathcal{S} (contrairement à l'opération de Steiner qui est définie sur 2^X tout entier).

Le livre de HARPER [98] et le “survey” de BEZRUKOV [28] donnent une généralisation de ces techniques et de ces résultats aux sous-graphes de produits cartésiens d'autres graphes réguliers, ainsi que des applications de ces résultats au problème arête-isopérimétrique. Pour les problèmes arête- ou sommet-isopérimétriques dans les graphes de Johnson (toujours ouverts à ce jour), AHLSSWEDE et CAI [6], ou encore DIEGO, SERRA et VENA [67], utilisent une opération assez naturelle de *décalage à gauche* (“pushing to the left”, “switching”, ou bien “shifting”) définie comme suit. Soit \mathcal{S} une famille de sous-ensembles de cardinal r d'un univers X à m éléments. Étant donné un ordre total arbitraire e_1, \dots, e_m des éléments de X et deux éléments $e_i < e_j$, l'opération de décalage à gauche de \mathcal{S} par rapport à ces deux éléments consiste à remplacer chaque ensemble S de \mathcal{S} contenant e_j mais pas e_i par $S \setminus \{e_j\} \cup \{e_i\}$ si ce dernier n'appartient pas déjà à \mathcal{S} . Cette opération préserve clairement la taille de \mathcal{S} et le cardinal r de ses ensembles, et nous pouvons aussi montrer qu'elle ne diminue pas le nombre d'arêtes du sous-graphe de Johnson induit par \mathcal{S} . Cependant, elle n'apporte pas suffisamment de structure au graphe résultant pour étudier (borner) sa dégénérescence (sa densité).

2.5 Généralisations et extensions de VC-dimension

L'extension de l'inégalité $\frac{|E|}{|V|} \leq \text{VC-dim}(\mathcal{S})$ à des classes de graphes plus générales (en étendant la notion de VC-dimension) constitue un problème important et intéressant. Dans les chapitres 5 et 6, nous présentons une généralisation de ce résultat de densité aux sous-graphes de produits cartésiens de graphes et aux sous-graphes de demi-cubes, respectivement. Nous présentons ici d'autres travaux concernant ce genre de généralisation.

Puisque nous avons parlé jusqu'ici de classification binaire, une généralisation naturelle de la VC-dimension concerne donc une extension aux prédictions avec un nombre $k \in \mathbb{N}^*$ de classes. C'est-à-dire que l'ensemble Y vaut alors $\{1, \dots, k\}$ et les classes de concepts sont donc des sous-ensembles de $\{1, \dots, k\}^n$ (toujours avec n le cardinal du domaine X). Le graphe induit par cette classe n'est donc plus un sous-graphe d'hypercube mais de graphe de Hamming. Ainsi, la classe des graphes de Hamming est bien présente dans la littérature. En particulier, des généralisations de pulvérisation et VC-dimensions ont été utilisées dans le cas des graphes de Hamming (voir *Convergence of Stochastic Processes* [133] ou « On learning sets and functions » [123]), conduisant à des résultats de densité par RUBINSTEIN, BARTLETT et RUBINSTEIN [137], et un lemme de Sauer étendu par HAUSSLER et LONG [101]. CESA-BIANCHI et HAUSSLER [41], quant à eux ont proposé une généralisation du lemme de Sauer, plus axée sur l'aspect théorie des graphes, puisqu'ils considère non pas des graphes Hamming, mais des produits cartésiens $F^m = F \square \dots \square F$ de m fois un graphe non-orienté F .

HAUSSLER et WELZL [102] définissent une notion de VC-dimension d'un graphe $G = (V, E)$ comme la VC-dimension d'une famille d'ensembles formée par les voisinages fermés des sommets de G . Ils démontrent alors que cette VC-dimension est d'au plus 4 pour les graphes planaires. ANTHONY, BRIGHTWELL et COOPER [17] montrent ensuite que cette VC-dimension est d'au plus d si G ne contient pas de mineur K_{d+1} (ils étudient aussi cette notion de VC-dimension pour les graphes aléatoires). Ces deux résultats ont par la suite été étendus aux familles de boules fermées de rayon fixé par CHEPOI, ESTELLON et VAXÈS [55, Proposition 1 & Remarque 1]. KRANAKIS et al. [111] introduisent des notions de VC-dimensions pour d'autres familles d'ensembles naturelles sur les graphes : des familles d'arbres induits, de sous-graphes connexes, de chemins, de cliques, d'étoiles, etc. Ils étudient les problèmes de complexité liés au calcul de ces VC-dimensions et donnent des bornes supérieures (dépendantes de paramètres de graphes) pour certaines d'entre elles. Pour améliorer les bornes sur le temps des requêtes d'un algorithme de plus courts chemins sur des réseaux réels, ABRAHAM et al. [1] exploitent un concept de VC-dimension pour des familles de plus courts chemins dans des graphes où ces derniers sont uniques.

D'autres notions de VC-dimensions sont définies par la pulvérisation, non pas de tous les sous-ensembles d'un ensemble, mais seulement de certains d'entre eux. Par exemple, BOUSQUET et THOMASSÉ [36] ont défini une notion de VC-dimension (la 2VC-dimension) par rapport à la pulvérisation des sous-ensembles de cardinal 2 d'un ensemble. Avec cela, ils établissent la propriété d'Erdős-Posá pour la famille des boules de rayon r dans les graphes de 2VC-dimension bornée.

Chapitre 3

Schémas d'étiquetage

Sommaire

3.1	Introduction	45
3.2	Schémas d'adjacence	46
3.2.1	Arboricité et densité	47
3.2.2	Graphes universels	49
3.3	Schémas de distance	51
3.3.1	Distances dans les arbres	53
3.4	Schémas de routage	54
3.4.1	Routage dans les arbres	55

Dans ce chapitre, nous introduisons la notion de schéma d'étiquetage. Nous nous intéressons plus spécifiquement aux schémas d'étiquetage permettant de répondre à des requêtes d'adjacence (section 3.2), de distance (section 3.3) ou de routage (section 3.4).

3.1 Introduction

Les matrices d'adjacence et les listes d'adjacence sont deux façons classiques de représenter un graphe $G = (V, E)$ en mémoire. La première permet de répondre à l'adjacence de deux sommets en temps constant pour un stockage en mémoire quadratique par rapport au nombre de sommets. La seconde nécessite un moindre espace de stockage (linéaire en $|E|$) en contrepartie de requêtes d'adjacence moins efficaces (linéaires $|V|$). Quoiqu'il en soit, ces deux représentations ne sont pas particulièrement adaptées au stockage de grands graphes car, pour répondre à des requêtes (d'adjacence, distance, routage, etc.) sur le graphe G , il est nécessaire d'accéder à toute la structure de donnée. Un schéma d'étiquetage est un procédé permettant de distribuer la représentation de G sur ses sommets de telle sorte qu'un type de requête (à choisir au préalable) puisse être

satisfait “rapidement” avec quelques informations locales, et sans connaissance supplémentaire sur le graphe.

Plus formellement, un *schéma d'étiquetage* sur une famille de graphes \mathcal{G} consiste en une fonction d'encodage $C_G : V(G) \rightarrow \{0, 1\}^*$, et en une fonction de décodage. La fonction d'encodage distribue des étiquettes binaires à chaque sommet d'un graphe $G \in \mathcal{G}$ avec une connaissance totale dudit graphe. La fonction de décodage est alors capable, étant données les étiquettes d'un nombre restreint de sommets (données par la fonction précédente) et avec la seule connaissance de ces étiquettes, de répondre à certaines requêtes pré-établies. Ces requêtes pré-établies peuvent être de nature assez variées, par exemple : adjacence ou distance entre deux sommets ; routage d'un sommet source vers un sommet cible ; ancêtre commun dans un arbre, etc. Évidemment, les fonctions d'encodage et de décodage dépendent fortement des requêtes auxquelles nous souhaitons répondre.

Remarque 7. Il n'est pas toujours précisé dans la définition d'un schéma d'étiquetage que les étiquettes doivent jouer un rôle d'identifiant pour les sommets. Pour notre part, nous supposerons toujours ces étiquettes uniques.

Notons qu'avec l'ajout à la définition de schéma de la remarque précédente, $\Omega(\log n)$ sera toujours une borne inférieure pour une famille de graphes avec n sommets (car $\lceil \log n \rceil$ bits sont nécessaires pour encoder n entiers distincts).

Un problème important dans ce cadre consiste à trouver des classes de graphes naturelles admettant des schémas d'étiquetage (d'adjacence, distance, routage, etc.) efficaces. C'est-à-dire, avec des étiquettes de taille poly-logarithmiques, décodables en temps poly-logarithmique. En ce qui concerne l'étude de l'efficacité du décodage (ou de l'encodage), il est nécessaire de définir un modèle de calcul (c.-à-d., les opérations que l'on estime “lentes” et celles que nous estimons “rapides”). Le modèle communément considéré est le *modèle RAM* (pour “*Random Access Machine*”), faisant des hypothèses raisonnables relativement au fonctionnement d'un ordinateur classique. Dans ce modèle, les opérations standards d'arithmétique, de transfert de données ou de contrôle (telles qu'addition, multiplication, comparaisons, modulo, partie entière, branchements conditionnels, appels de sous-routines, lecture, copie, etc.) sur des mots de $O(\log n)$ bits sont supposées effectuées en temps constant.

3.2 Schémas d'adjacence

Dans un *schéma d'étiquetage d'adjacence* (ou ALS pour “*adjacency labeling scheme*”) pour une famille de graphes \mathcal{G} , la fonction de décodage $D_G : \{0, 1\}^* \times$

$\{0, 1\}^* \rightarrow \{0, 1\}$ a accès aux étiquettes binaires de deux sommets de G et retournera 1 si et seulement si les sommets qu'elles encodent sont adjacents.

Les premiers schémas d'étiquetage étudiés s'intéressaient aux requêtes d'adjacence. BREUER [38], notamment, a prouvé que les sommets de tout graphe G peuvent être étiquetés de sorte que leur adjacence (et leur distance) puisse être déduite de la distance de Hamming entre leurs étiquettes. Une telle représentation peut cependant utiliser des étiquettes particulièrement longues. Par la suite, ce problème est étudié à nouveau sous le nom de *schéma d'étiquetage d'adjacence* ou de *représentation implicite de graphe* par MULLER [120] et KANNAN, NAOR et RUDICH [106]. Leur but est alors de rendre ces représentations “efficaces”. En effet, pour que le fait de distribuer la représentation d'un graphe sur ses sommets présente un réel intérêt, les auteurs demandent que les informations locales stockées soient courtes et décodables en temps raisonnable. En fait, KANNAN, NAOR et RUDICH [106] ajoutent à la définition d'un schéma d'étiquetage que les étiquettes doivent être de longueur poly-logarithmiques en le nombre de sommets du graphe, et que leur décodage doit s'effectuer en temps polynomial en cette longueur (c.-à-d., en temps poly-logarithmique en la taille du graphe). Cette définition d'efficacité semble raisonnable, car tous les graphes n'admettent pas d'étiquetage sous ces conditions.

Proposition 5 (KANNAN, NAOR et RUDICH [106]). *Une famille de graphes sur n sommets dont le cardinal est supérieur à $2^{\Omega(n \log n)}$ n'admet pas de schéma d'adjacence efficace.*

Précisons que, dans leur définition, KANNAN, NAOR et RUDICH, supposent aussi que les étiquettes attribuées sont uniques. Notons aussi qu'en conséquence de la proposition 5, les graphes biparties et les graphes cordaux, en particulier, n'admettent pas de schéma efficace.

3.2.1 Arboricité et densité

Nous continuons avec la notion d'arboricité et son lien avec la densité des graphes (au sens où nous l'avons définie dans la section 1.1.3). Comme nous le verrons, cette notion est assez centrale dans la plupart des résultats sur les schémas d'adjacence et des graphes universels (induits) que nous aborderons ensuite.

L'*arboricité* $a(G)$ d'un graphe G correspond au nombre minimum de forêts nécessaires pour couvrir toutes les arêtes de G . Le théorème classique de NASH-WILLIAMS [121] établit que les graphes de faible arboricité sont “creux” (“*sparse*”). C'est-à-dire qu'ils ne possèdent pas de sous-graphes “denses”. Autrement dit, l'arboricité d'un graphe est étroitement liée à sa densité.

Théorème 9 (NASH-WILLIAMS [121]). *Les arêtes d'un graphe $G = (V, E)$ peuvent être partitionnées en k forêts si et seulement si $|E(H)| \leq k(|V(H)| - 1)$ pour tout sous-graphe H de G . C'est-à-dire,*

$$a(G) = \max_{H \subseteq G} \left\lceil \frac{|E(H)|}{|V(H)| - 1} \right\rceil$$

Remarque 8. Bien que le nombre de sous-graphes d'un graphe quelconque G soit exponentiel, le calcul de l'arboricité de G peut se faire en temps polynomial par un résultat de GABOW et WESTERMANN [77].

KANNAN, NAOR et RUDICH [106] ont démontré la proposition suivante :

Proposition 6. *La famille de graphes avec n sommets d'arboricité k admet un schéma d'étiquetage d'adjacence utilisant $(k + 1) \lceil \log n \rceil$ bits.*

Démonstration. En effet, la fonction d'encodage d'un schéma d'adjacence sur \mathcal{G} attribuera à chaque sommet d'un graphe $G \in \mathcal{G}$ un identifiant unique (de $O(\log n)$ bits) et y concatènera celui de son père dans chacune des k forêts nécessaires à la couverture des arêtes de G . Étant données deux étiquettes, la fonction d'encodage retournera 1 si et seulement si l'identifiant d'un des sommets apparaît dans l'ensemble des pères de l'autre (voir la figure 3.1). \square

Assez récemment, ALSTRUP, DAHLGAARD et KNUDSEN [9] ont proposé un schéma d'adjacence pour la famille des forêts à n sommets utilisant des étiquettes de $\lceil \log n \rceil + O(1)$ bits, et autorisant un décodage en temps constant. Ce résultat est donc optimal (à constante additive près) et il clôt, en quelque sorte, une série de papiers (CHUNG [62] et ALSTRUP et RAUHE [15], entre autres) atteignant des longueurs d'étiquettes de $\log n + O(\log \log n)$ bits, puis $\log n + O(\log^* n)$ bits.

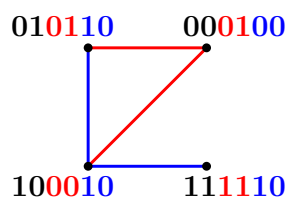


FIGURE 3.1 – Étiquetage d'adjacence d'un graphe dont les arêtes sont couvertes par deux forêts. La partie noire d'une étiquette correspond à l'identifiant unique du sommet, les parties rouge et bleue correspondent respectivement à l'identifiant de son père dans les forêts rouge (de racines 01 et 11) et bleue (de racines 00 et 10).

Ainsi, en utilisant la proposition 6 de KANNAN, NAOR et RUDICH [106], borner la densité (et donc l'arboricité) d'un graphe fourni un bon moyen obtenir

des bornes supérieures sur la longueur des étiquettes attribuées par un schéma d'adjacence. C'est ce qui a initialement motivé les travaux que nous présentons dans la partie II. Dans certain cas, cette borne supérieure peut être largement sous-optimale. Par exemple, les graphes complets sont d'arboricité maximale, pourtant un schéma d'adjacence est trivial à établir pour eux. Pour autant, des schémas basés sur une couverture par les arbres peuvent donner des résultats optimaux (à constante multiplicative près) ou, à défaut, les meilleurs que l'on soit parvenu à obtenir. C'est notamment le cas pour les graphes planaires (dont l'arboricité est inférieure ou égale à 3, en conséquence assez directe de la formule d'Euler¹), des graphes avec mineur exclu, ou des sous-graphes d'hypercubes (dont l'arboricité est logarithmique, comme nous l'avons vu dans le chapitre 1). Plus exactement, le meilleur schéma d'adjacence actuel pour les graphes de mineur exclu fixé (dont les graphes planaires sont un exemple) est dû à GAVOILLE et LABOUREL [81]. Il est basé, non pas sur une décomposition en forêt, mais sur une décomposition en graphes de largeur arborescente bornée, dont les auteurs montrent qu'ils admettent un schéma d'adjacence utilisant des étiquettes de $(1 + o(1)) \log n$ bits. Les graphes planaires, en particulier, sont décomposables en deux tels graphes.

3.2.2 Graphes universels

La notion de graphe universel, introduite par RADO [135] en 1964, est la suivante. Un graphe U est *universel* pour une famille de graphes \mathcal{G} (ou \mathcal{G} -universel) si tout graphe $G \in \mathcal{G}$ est isomorphe à un sous-graphe de U . Nous pouvons remarquer qu'une clique avec n sommets est un graphe universel pour la famille des graphes avec n sommets. Cependant, le problème intéressant consiste à trouver, pour une famille de graphes \mathcal{G} , un graphe \mathcal{G} -universel U avec un nombre minimal d'arêtes. L'une des motivations originelles justifiant l'intérêt des graphes universels réside dans la conception de circuits intégrés (voir, par exemple, les travaux de VALIANT [147] ou BHATT et LEISERSON [29]).

Un graphe U est *\mathcal{G} -universel induit* s'il contient tous les graphes de \mathcal{G} en tant que sous-graphes induits (voir la figure 3.2). Dans le cas des graphes universels induits, le but est alors de minimiser le nombre de sommets.

Dans « Universal graphs and induced-universal graphs » [62], CHUNG établit des liens entre graphe universel "simple" et son pendant induit. Elle donne, en particulier, des bornes sur le nombre d'arêtes et de sommets d'un graphe \mathcal{G} -universel induit, par rapport à ceux d'un graphe \mathcal{G} -universel (où \mathcal{G} est une famille de graphes quelconques). De ces résultats, elle déduit des bornes inférieures et supérieures sur le nombre de sommets d'un graphe universel induit

1. $\#sommets - \#arêtes + \#faces = 2$ pour tout graphe planaire

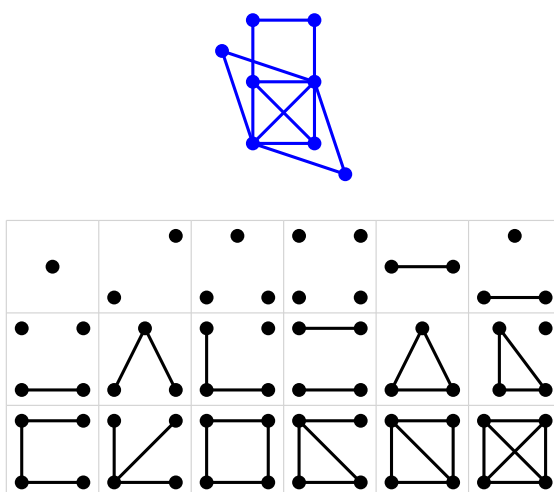


FIGURE 3.2 – Exemple d'un graphe universel induit (en bleu) pour la famille des graphes planaires avec au plus 4 sommets (graphes noirs).

pour des familles de graphes d'arboricité bornée, dont les arbres et les graphes planaires sont d'importants représentants.

Dans « Implicit representation of graphs » [106], KANNAN, NAOR et RUDICH remarquent que les notions de schéma d'adjacence et de graphe universel induit sont étroitement liées. En effet, une famille de graphes \mathcal{G} admet un schéma d'adjacence utilisant des étiquettes de k bits si et seulement s'il existe un graphe \mathcal{G} -universel induit possédant 2^k sommets. Sachant cela, SLEATOR et TARJAN [142] déduisent de leur schéma d'adjacence optimal pour les forêts une borne supérieure de $O(n)$ nœuds pour le graphe universel induit des forêts de n nœuds. Cette borne supérieure améliore leur précédent résultat (présenté dans « Small induced-universal graphs and compact implicit graph representations » [15]) de $n2^{O(\log^* n)}$, et elle atteint la borne inférieure (fournie dans ce même papier précédent) de $\Omega(n)$. Il en découle que la famille des graphes à n sommets, dont les arêtes peuvent être couvertes par k forêts, admet un graphe universel induit avec $\Theta(n^k)$ sommets. Une partie des résultats optimaux actuels concernant les graphes universels induits pour les familles des graphes généraux, bipartis ou planaires, entre autres, sont rappelés dans « Adjacency Labeling Schemes and Induced-Universal Graphs » [14]. Ce papier est toutefois consacré à la présentation d'un étiquetage d'adjacence des graphes (non-orientés) généraux avec n sommets utilisant des étiquettes de $n/2 + O(1)$ bits. Les auteurs en déduisent que la famille des graphes généraux de n sommets admet un graphe universel induit avec $O(2^{n/2})$ sommets, ce qui est asymptotiquement optimal. En effet, une borne inférieure de $2^{\frac{n-1}{2}}$ pour cette famille de graphes est facile à obtenir puisque le nombre de graphes (différents) avec n sommets est d'au moins $2^{\binom{n}{2}}/n!$ (ce genre d'argument de comptage est souvent utilisé pour obtenir des bornes inférieures).

Avec leur résultat, ALSTRUP et al. [14] améliorent la précédente borne supérieure de $n/2 + O(\log n)$, obtenue par MOON [117] près de 50 années auparavant.

3.3 Schémas de distance

Dans un schéma d'étiquetage de distance (ou DLS pour “Distance Labeling Scheme”) sur une famille de graphes \mathcal{G} , la fonction de décodage $D'_G : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \mathbb{N}$ a accès aux étiquettes de deux sommets u et v de G et calcule la distance $d_G(u, v)$.

La notion de schéma d'étiquetage de distance a été introduite pour la première fois dans une série de papiers de PELEG [131, 132] et GAVOILLE et al. [85]. Avant cela, des notions assez proches existaient déjà avec d'autres vocabulaires, moins unifiés. Par exemple, le problème du plongement dans les cubes écrasés (“squashed cubes”), introduit par WINKLER [151], était équivalent à un schéma de distance avec des étiquettes de longueur $\log n$ fois la dimension du cube.

L'un des résultats principaux sur les DLS, dû en particulier à WINKLER [151], GAVOILLE et al. [85] ou ALSTRUP et al. [11], est le fait que la famille des graphes généraux sur n sommets admette un schéma de distance avec des étiquettes de $O(n)$ bits. Ce schéma est asymptotiquement optimal puisqu'un simple argument de comptage de ces graphes suffit à montrer que $\Omega(n)$ bits sont nécessaires. Un autre résultat intéressant est qu'il existe un schéma de distance pour les arbres avec des étiquettes de $O(\log^2 n)$ bits. De tels schémas ont été proposés par PELEG [131] et ALSTRUP et al. [13], et plus récemment par FREEDMAN et al. [75] (qui obtiennent des requêtes en temps constant). Plusieurs classes de graphes contenant les arbres possèdent aussi de tels schémas de distance avec étiquettes de $O(\log^2 n)$ bits. C'est le cas notamment des graphes de largeur arborescente bornée (GAVOILLE et al. [85]), des graphes distance-héréditaires (GAVOILLE et PAUL [83]), des graphes de largeur de clique bornée (COURCELLE et VANICAT [64]) et des graphes de courbure négative ou nulle (CHEPOI, DRAGAN et VAXÈS [53]). GAVOILLE et al. [85] et ALSTRUP et al. [13] ont aussi donné une borne inférieure de $\Omega(\log^2 n)$ bits sur la taille des étiquettes pour les arbres, ce qui implique que tous les résultats mentionnés juste avant sont aussi optimaux. D'autres familles de graphes ont été étudiées par BAZZARO et GAVOILLE [27] et GAVOILLE et PAUL [84], entre autres, telles que les graphes d'intervalles, les graphes de permutations ainsi que leurs généralisations (pour lesquelles il existe une borne optimale de $\Theta(\log n)$ bit), et les graphes planaires (pour lesquelles GAVOILLE et al. [85] donnent une borne inférieure de $\Omega(n^{\frac{1}{3}})$ bits, et GAWRYCHOWSKI et UZNANSKI [87] fournissent une borne supérieure de $O(\sqrt{n})$ bits).

Il existe aussi des résultats concernant des schémas d'étiquetage de distance

approximée, c'est-à-dire des schémas dont la fonction de décodage retourne la distance entre les sommets avec une erreur additive et/ou multiplicative. Pour les graphes arbitraires, le résultat le plus important, donné par THORUP et ZWICK [146], est un schéma de distance avec erreur multiplicative $(2k - 1)$ et tailles d'étiquettes $O(n^{1/k} \log^2 n)$ pour tout choix d'un entier $k \geq 1$. Dans « Approximate distance labeling schemes » [79], il est démontré que les arbres (et les graphes de largeur arborescente bornée en général) admettent un DLS avec erreur multiplicative $1 + 1/\log n$ et taille d'étiquettes $O(\log n \cdot \log \log n)$. Ce résultat est serré en terme de taille d'étiquette et d'approximation. Les mêmes auteurs conçoivent aussi un DLS avec erreur additive $O(1)$ et étiquettes $O(\log^2 n)$ pour plusieurs autres familles de graphes telles que les graphes de plus long cycle induit borné ou, plus généralement, de longueur arborescente bornée. Pour les graphes δ -hyperboliques, GAVOILLE et LY [82] et CHEPOI et al. [52] construisent des schémas de distance avec erreur additive $O(\delta \log n)$ et étiquettes de longueurs $O(\log^2 n)$. Enfin, certains résultats concernent des approximations affines qui combinent erreur additive et multiplicative (par exemple, les travaux de ABRAHAM et GAVOILLE [3]). Nous pouvons remarquer que les graphes de longueur arborescente (“*treelength*”) bornée ont, de fait, une hyperbolicité bornée et, plus important, peuvent être plongés dans des arbres avec une distorsion bornée, dépendante de la longueur arborescente. Cela donne une vue alternative du précédent résultat de GAVOILLE et al. [79]. Nous pouvons aussi noter que tout DLS exact pour ces familles de graphes doit attribuer des étiquettes de $\Omega(n)$ bits dans le pire des cas. Cela peut être expliqué par le fait que des propriétés telles que l'hyperbolicité, la longueur arborescente, ou la quasi-isométrie avec un arbre sont des propriétés globales. Elles permettent donc un comportement local assez arbitraire, ce qui conduit à des erreurs arbitrairement grandes lorsque les requêtes portent sur des sommets suffisamment proches.

Une alternative au calcul de distances approchées consiste à donner des distances exactes seulement pour certaines paires de sommets. L'essentiel des résultats sur cette question concerne le calcul exact de distance entre des sommets “éloignés” (c.-à-d., à distance au moins D l'un de l'autre) ou bien “proches” (c.-à-d., à distance au plus D , pour un certain seuil D). Par exemple, BOLLOBÁS, COPPERSMITH et ELKIN [32] introduisent la notion de DLS D -préservant, qui consiste en un DLS ne calculant des distances exactes qu'entre les sommets à distance au plus D l'un de l'autre. Le DLS qu'ils présentent (pour des graphes quelconques) utilise des étiquettes de longueur $O(\frac{n}{D} \log^2 n)$. Ce résultat a ensuite été amélioré en $O(\frac{n}{D} \log^2 D)$ par ALSTRUP et al. [10], qui montrent aussi une borne inférieure de $\Omega(\frac{n}{D})$. Enfin, la meilleure borne actuelle de $O(\log^2 D + \frac{n}{D} \log D \log \log D)$ est obtenue par KOSOWSKI et VIENNOT [110]. Pour les arbres enracinés, KAPLAN et MILO [107] introduisent la notion de DLS pour les distances “courtes”. Plus précisément, leur DLS calcule la distance au plus proche ancêtre commun (c.-à-d., le DLS calcule la distance exacte pour les sommets à distance au plus D).

La meilleure borne supérieure connue pour la taille des étiquettes dans un tel schéma est de $\log n + O(D \log(D \log(n/D)))$, obtenue par GAVOILLE et LABOUREL [80]. Lorsque $D \geq 2$, ALSTRUP, BILLE et RAUHE [8] donnent aussi une borne inférieure de $\log n + \Omega(\log \log n)$.

3.3.1 Distances dans les arbres

Nous présentons ici le schéma de distance (`Enc_Dist_Arbre`, `Dist_Arbre`) pour les arbres proposé par PELEG [131]. Bien que ce schéma ne soit pas optimal en terme de rapidité des requêtes, il a l'avantage d'être assez simple. De plus, il constitue une introduction aux idées présentées dans les chapitres 7 et 8 qui peuvent être considérées comme des généralisations de celles à la base du schéma de PELEG. Notons tout de même que, lorsque nous intégrerons (dans la partie III) des schémas de distance pour les arbres à nos propres schémas, nous considérerons celui, plus efficace, de FREEDMAN et al. [75].

Pour un sommet $u \in V$ d'un graphe $G = (V, E)$, posons $M(u) := \sum_{v \in V} d_G(u, v)$. Un sommet m minimisant cette fonction M est appelé un *sommet médian* de G . La proposition suivante est bien connue :

Proposition 7. *Tout arbre T possède soit un sommet médian, soit deux reliés par une arête. De plus, v est un sommet médian d'un arbre T si et seulement si chaque sous-arbre induit par $V(T) \setminus \{v\}$ contient au plus la moitié des sommets de T .*

Du fait de la proposition 7 précédente, un sommet médian dans un arbre est plus communément nommé un *centroïde* (ou *centre*, comme initialement introduit par JORDAN [105]).

La procédure `Enc_Dist_Arbre` donne une étiquette $LD_T(v)$ à tout sommet v de l'arbre T de la façon suivante :

- (1) Donner à chaque sommet v un identifiant unique $\text{id}(v)$;
- (2) Trouver un sommet médian m de T ;
- (3) Pour tout sommet v de T , concaténer $(\text{id}(m), d_T(v, m))$ à l'étiquette courante de v ;
- (4) Répéter à l'étape 2 pour chaque sous-arbre d'au moins deux sommets créé par la suppression de m .

Étant données les étiquettes $LD_T(u)$ et $LD_T(v)$, la procédure `Dist_Arbre` peut trouver le dernier séparateur commun m à u et v (c.-à-d., le dernier médian commun stocké par u et v dans leurs étiquettes) et retourner $d_T(u, m) + d_T(v, m)$ en tant que la distance $d_T(u, v)$ (voir la figure 3.3).

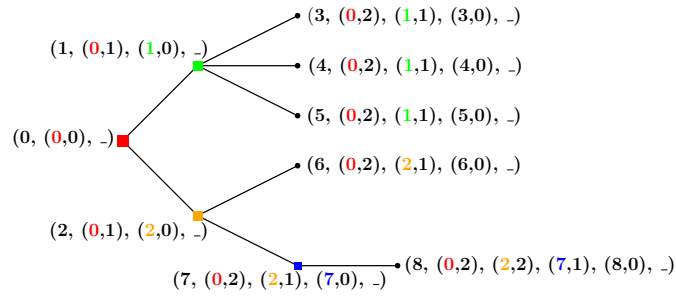


FIGURE 3.3 – Exemple de l'étiquetage d'un arbre avec le schéma de distance de PELEG.

3.4 Schémas de routage

Dans un *schéma d'étiquetage de routage* (ou RLS pour “*Routing Labeling Scheme*”), la fonction de décodage $D_G'' : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \mathbb{N}$ a accès à l'étiquette d'un sommet *source* u et à celle d'un sommet *cible* v , et elle calcule un port $\text{port}(u, v)$ à prendre depuis u pour se rapprocher de v . Rappelons qu'un port $\text{port}(u, v)$ est un identifiant unique donné à une arête $uv \in E$ relativement au sommet $u \in V$. Plus généralement, si u et v sont deux sommets quelconques de G , $\text{port}(u, v)$ dénote n'importe quelle valeur $\text{port}(u, v')$ telle que $uv' \in E$ et $v' \in I(u, v)$ (par convention, nous pourrions poser $\text{port}(u, u) := 0$). Un graphe dans lequel chaque arête possède un port relativement à chacune de ses extrémités est dit *avec ports*. Ce sera le cas de tous les graphes que nous considérerons lorsque nous nous intéresserons à des problèmes de routage.

Router de l'information est l'une des tâches principales qu'un réseau distribué doit être capable d'effectuer. De ce fait, la conception de schémas de routage efficaces est un sujet bien étudié. Le livre *Distributed Computing : A Locality-Sensitive Approach* [130], par exemple, donne un aperçu général des résultats sur ce domaine. Un moyen trivial pour obtenir un RLS exact (c.-à-d., un schéma routant sur des plus courts chemins) consiste à stocker dans chaque nœud une table de routage complète du réseau. Cette table indique, pour chaque destination, le port à emprunter depuis chaque source pour rejoindre la destination en passant par un plus court chemin. Cela donne un RLS exact avec des étiquettes de $O(n \log \Delta)$ bits, où Δ représente le degré maximum du graphe. Ce résultat, bien que peu satisfaisant est pourtant optimal dans le cas général par un résultat de GAVOILLE et PÉRENNÈS [86]. Pour les arbres, il existe des RLS exacts avec des étiquettes de longueur $(1 + o(1)) \log n$, construits par FRAIGNIAUD et GAVOILLE [74] et THORUP et ZWICK [145]. Des RLS exacts avec étiquettes de taille poly-logarithmiques existent aussi pour les graphes de largeur-arborescente, largeur de clique ou cordalité bornées (DRAGAN et YAN [70]), et pour les graphes

planaires de courbure non-positive (CHEPOI, DRAGAN et VAXÈS [53]). Pour la famille des graphes excluant un mineur fixé (graphes planaires ou de genre bornés entre autres), DRAGAN et YAN [70] donnent aussi un RLS exact utilisant des étiquettes de $O(\sqrt{n} \log^2 n / \log \log n)$ bits .

Puisque les RLS exacts pour des graphes généraux nécessitent des étiquettes de $O(n)$ bits, nous pouvons nous poser la question de RLS approchés, c'est-à-dire, de RLS ne routant pas les paquets sur des plus courts chemins mais sur des chemins faisant peu de détours, à la manière des RLS étudiés par COWEN [65], EILAM, GAVOILLE et PELEG [72] et THORUP et ZWICK [145]. Un RLS 3-multiplicatif (c.-à-d., avec erreur multiplicative 3) a été obtenu par COWEN [65]. Un RLS 5-multiplicatif avec étiquettes de taille $\tilde{O}(n^{1/2})$ a été donné par EILAM, GAVOILLE et PELEG [72]. THORUP et ZWICK [145] ont ensuite amélioré ces résultats en construisant un RLS $(4k - 5)$ -multiplicatif avec des étiquettes n'utilisant que $\tilde{O}(kn^{1/k})$ bits, pour tout $k \geq 2$. D'autres résultats concernent des RLS avec erreur affine comme, par exemple, ceux de ABRAHAM et GAVOILLE [3].

3.4.1 Routage dans les arbres

Le schéma de distance de la sous-section 3.3.1 précédente peut être adapté pour obtenir un schéma de routage. Pour la fonction d'encodage `Enc_Rout_Arbre`, il suffit simplement de remplacer $(\text{id}(m), d_T(u, m))$ à l'étage (3) par $(\text{id}(m), \text{port}(u, m), \text{port}(m, u))$. Alors, la fonction de décodage `Rout_Arbre` (u, v) retourne $\text{port}(u, m)$ (stocké dans l'étiquette de u) lorsque $m \neq u$, et $\text{port}(m, v)$ (stocké dans l'étiquette de v) sinon (ici encore, m dénote le dernier séparateur commun de u et v).

Chapitre 4

Graphes faiblement modulaires

Sommaire

4.1	Graphes faiblement modulaires	56
4.2	Graphes médians	58
4.2.1	Graphes médians et complexes cubiques $CAT(0)$	61
4.2.2	Problèmes de distances dans les graphes médians et les complexes cubiques $CAT(0)$	63
4.3	Graphes pontés	64
4.3.1	Graphes pontés et complexes systoliques	66
4.4	Courbure et théorème de Gauss-Bonnet	67

Dans ce chapitre, nous commençons par introduire une classe importante de la théorie métrique des graphes : la classe des graphes faiblement modulaires. Nous nous intéressons ensuite à deux de ses sous-classes, celles des graphes médians et pontés, car ce sont elles qui nous concerneront particulièrement dans la partie III. Nous rappelons quelques-unes de leurs nombreuses définitions/caractérisations et donnons une partie de leurs liens avec divers domaines. Nous évoquons, en particulier, les liens respectifs des graphes médians et pontés avec des complexes cubiques et simpliciaux, expliquant notamment leur importance en théorie géométrique des groupes (précisons que ces liens ne sont pas fondamentaux pour la compréhension du reste de ce manuscrit). En fin de chapitre, dans la section 4.4, nous rappelons les définitions utiles à l'énoncé du théorème de Gauss-Bonnet dans le cadre des graphes planaires. Ce résultat est loin du cœur de ce manuscrit, mais il nous sera utile dans le chapitre 8 (affirmation 54.1).

4.1 Graphes faiblement modulaires

Les graphes *faiblement modulaires* sont introduits par CHEPOI [47] et BANDELT et CHEPOI [22] comme les graphes satisfaisant la *condition du triangle (CT)* ainsi

que la *condition du quadrangle (CQ)*, décrites ci-dessous et illustrée en figure 4.1.

- (CQ) Pour tout quadruplet de sommets $u, v, w, z \in V$ tels que $d_G(u, v) = d_G(u, w) =: k$, $d_G(u, z) = k + 1$ et $v, w \sim z$, il existe un sommet $x \sim v, w$ tel que $d_G(u, x) = k - 1$.
- (CT) Pour tout triplet de sommets $u, v, w \in V$ tels que $d_G(u, v) = d_G(u, w) =: k$ et $u \sim v$, il existe un sommet $x \sim v, w$ tel que $d_G(u, x) = k - 1$.

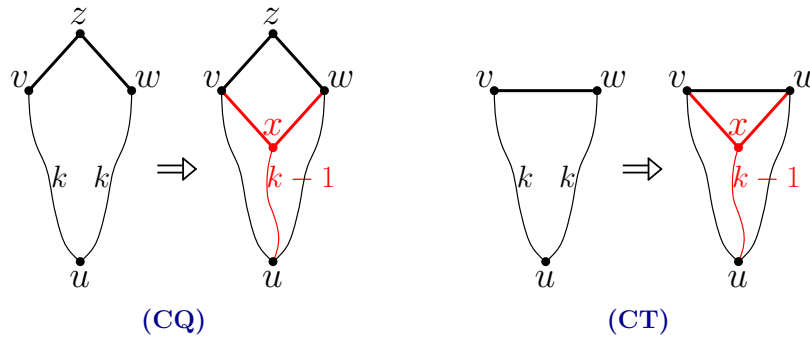


FIGURE 4.1 – Illustrations de la condition du triangle (à gauche) et du quadrangle (à droite).

Cette famille des graphes faiblement modulaires peut sembler assez *ad hoc* ; pour autant, elle contient un grand nombre de classes importantes de la théorie métrique des graphes, et elle les unifie par des caractérisations “local-vers-global” (et des considérations de courbure non-positive). Pour ce qui nous intéresse dans ce document, elle contient les graphes médians et les graphes pontés (dont nous parlons dans les deux sections suivantes). Parmi les sous-classes importantes des graphes faiblement modulaires, nous pouvons aussi citer les graphes modulaires et quasi-médians (deux généralisations des graphes médians), les graphes de Helly, ou encore les graphes “*dual polar*”. Pour des définitions et résultats sur ces familles, le lecteur pourra notamment consulter l’article « Weakly modular graphs and nonpositive curvature » [43].

Un *triangle métrique* dans un graphe connexe $G = (V, E)$ est un triplet de sommets (u_1, u_2, u_3) définissant trois intervalles qui s’intersectent deux-à-deux sur ces sommets uniquement. Plus précisément, u_1, u_2 et u_3 sont tels que $I(u_i, u_j) \cap I(u_j, u_k) = \{u_j\}$ pour tout $i \neq j \neq k$ de $\{1, 2, 3\}$. Un triangle métrique (u_1, u_2, u_3) est *équilatéral* si $d_G(u_1, u_2) = d_G(u_2, u_3) = d_G(u_1, u_3)$.

Remarque 9. Lorsque nous considérerons un triangle métrique (u_1, u_2, u_3) , ou le graphe induit par (u_1, u_2, u_3) , nous ferons bien souvent référence à l’enveloppe convexe du triplet (u_1, u_2, u_3) .

Soient x, y et z trois sommets d'un graphe connexe $G = (V, E)$. Un triangle métrique (x', y', z') tel que

$$\begin{cases} d_G(x, y) = d_G(x, x') + d_G(x', y') + d_G(y', y) \\ d_G(y, z) = d_G(y, y') + d_G(y', z') + d_G(z', z) \\ d_G(x, z) = d_G(x, x') + d_G(x', z') + d_G(z', z) \end{cases}$$

est appelé un *quasi-médian* de x, y et z . Nous pouvons remarquer que tout triplet de sommets d'un graphe connexe admet un quasi-médian. Les graphes faiblement modulaires sont caractérisés par la propriété forte suivante de leurs triangles métriques (c.-à-d., quasi-médians) :

Théorème 10 (CHEPOI [47]). *Un graphe $G = (V, E)$ est faiblement modulaire si et seulement si, pour tout triangle métrique (u_1, u_2, u_3) de G , et pour tout $x, y \in I(u_1, u_2)$, l'égalité $d_G(u_3, x) = d_G(u_3, y)$ est vérifiée. Il s'en suit que chaque triangle métrique d'un graphe faiblement modulaire est équilatéral.*

4.2 Graphes médians

Les graphes médians et les structures médianes reliées (telles que les algèbres médianes ou les complexes médians), ont une littérature très fournie. Le concept de graphe médian est apparu bien avant dans le cadre des treillis distributifs (BIRKHOFF et KISS [31] et AVANN [18]), mais le terme de graphe médian n'a été introduit que plus tard par NEBESKÝ [124]. Ces structures ont été étudiées dans des contextes différents par un grand nombre d'auteurs depuis plus d'un demi-siècle. Les structures médianes sont régulièrement redécouvertes sous divers noms, et dans des domaines variés. Elles sont l'objet de nombreux "surveys" recensant leurs nombreuses caractérisations et propriétés (BANDELT et CHEPOI [23], KLAVŽAR et MULDER [108] et KNUTH [109]).

Un graphe G est *médian* si l'intersection $I(u, v) \cap I(v, w) \cap I(w, u)$ est un singleton pour chaque triplet u, v, w de ses sommets. L'unique sommet $\bar{m}(u, v, w) \in I(u, v) \cap I(v, w) \cap I(w, u)$ est alors nommé le *médian* de u, v et w .

Nous commençons par énoncer les propriétés des graphes médians qui nous fourniront les outils utiles à certaines preuves du chapitre 7.

Lemme 5. *Tout graphe médian $G = (V, E)$ satisfait la condition du quadrangle.*

Démonstration. Soit u, v, w et z quatre sommets de G tels que $z \sim u, v$, $d_G(u, w) = d_G(v, w) =: k$ et $d_G(w, z) = k + 1$. Soit x le médian du triplet u, v, w . Alors x doit être adjacent à v et w . En effet, G ne contient aucun triangle, et v et w sont à distance 2 (par l'existence de $z \sim v, w$). Aussi, $d_G(u, x)$ est forcément égale à $k - 1$

car $x \in I(u, v) \cap I(u, w)$. Cela conclut la preuve. Nous pouvons aussi remarquer que x est unique car tout sommet x' adjacent à v et w et étant à distance $k - 1$ de u doit être un médian de u, v et w , impliquant $x' = x$. \square

Remarquons que les graphes médians satisfont aussi la condition du triangle dans le sens où aucun triplet de sommets d'un graphe médian ne se place dans ses conditions. Il s'en suit que les graphes médians forment une sous-classe des graphes faiblement modulaires. Plus précisément, ils sont une sous-classe d'une autre famille de graphes faiblement modulaires sans triangles : les *graphes modulaires* (c.-à-d., graphes dans lequel chaque triplet de sommet admet au moins un sommet médian). En fait, les graphes médians sont aussi caractérisés comme suit :

Théorème 11 (BANDELT [20]). *Un graphe est médian si et seulement il est modulaire et ne contient pas de $K_{2,3}$ induit (c.-à-d., G est biparti, satisfait (CQ) et ne contient pas de $K_{2,3}$ induits).*

Le résultat suivant est établi par CHEPOI [47] et découle de la caractérisation “local-vers-global” de la convexité et de la “gatedness” dans les graphes faiblement modulaires. Il peut donc être formulé dans le cadre général des graphes faiblement modulaires. Nous le formulons et le prouvons cependant dans le contexte plus restreint des graphes médians, car c'est dans ce cadre que nous utiliserons cette propriété dans le chapitre 7.

Lemme 6. *Pour un graphe médian $G = (V, E)$ et un sous-ensemble de sommets $A \subseteq V$, les propriétés suivantes sont équivalentes :*

- (i) $H := G[A]$ est connexe et A est localement convexe (c.-à-d., $x, y \in A$ et $d_G(x, y) = 2 \Rightarrow I(x, y) \subseteq A$);
- (ii) A est convexe ;
- (iii) A est porté.

Démonstration. (i) \Rightarrow (ii) : Soit u et v deux sommets quelconques de A . Nous montrons que $I(u, v) \subseteq A$ par induction sur la distance $d_H(u, v)$ entre u et v dans H . Si $d_H(u, v) = 2$, alors la propriété est vérifiée du fait de la convexité locale de A . Considérons que $k := d_H(u, v) \geq 3$ et supposons que $I(u', v') \subseteq A$ pour toute paire de sommets $(u', v') \in A^2$ à distance $k - 1$ l'un de l'autre. Nous montrons que la propriété reste vraie pour les sommets à distance k l'un de l'autre par l'absurde. Supposons qu'il existe sommet $x \in V \setminus A$ sur un plus court chemin P entre u et v . Remarquons que, comme x n'appartient pas à H , le voisin u' de u sur P ne peut pas non plus appartenir à H puisqu'il est à distance $k - 1$ de v . Considérons un sommet u'' voisin de u dans $A \cap I(u, v)$ et le sommet médian $m := \bar{m}(u', u'', v)$. Puisque $m \in I(u'', v)$, $m \in A$ et comme $m \in I(u', u'')$, $d_G(u, m) = 2$. Mais alors u' appartient à $I(u, m)$, ce qui contredit son choix hors de A .

(ii) \Rightarrow (iii) : Supposons, par l'absurde, que A soit convexe mais pas porté. Alors il existe un sommet $u \in V \setminus A$ qui ne possède pas de porte sur A . Soit x un sommet de A parmi les plus proches de u . Puisque u n'est pas la porte de u sur A , il existe un sommet $y \in A$ tel que $x \notin I(u, y)$. Soit m le médian du triplet (u, x, y) . Puisque $x \notin I(u, y)$, $m \neq x$, et comme $m \in I(x, y)$ et H est convexe, m appartient à A . Mais puisque $m \in I(x, u)$ et $m \neq x$, $d_G(u, m) < d_G(u, x)$, contredisant le choix de x .

(iii) \Rightarrow (i) : Tout ensemble porté A induit un sous-graphe connexe. Pour montrer que A est localement convexe, choisissons deux sommets $x, y \in A$ avec $d_G(x, y) = 2$ et avec un voisin commun u . Si $u \notin A$, alors u n'a clairement pas de porte sur A car $I(u, x) \cap A = \{x\}$ et $I(u, y) \cap A = \{y\}$. \square

Bien que le lien ne soit pas évident de par leur définition, les graphes médians sont intimement liés aux hypercubes (voir l'illustration 4.2 d'un graphe médian quelconque). Les graphes médians sont bipartis. Les arbres, les hypercubes, les grilles rectangulaires ou les diagrammes de Hasse de treillis distributifs en sont des exemples courants (voir, par exemple, le "survey" de BANDELT et CHEPOI [23]). En fait, les graphes médians peuvent être obtenus par amalgames d'hypercubes, et ils sont des sous-graphes isométriques d'hypercubes (BANDELT et DE VEL [25] et MULDER [118]). Plus précisément, BANDELT [19] a démontré que *les graphes médians sont exactement les retractés d'hypercubes* (voir la figure 1.4). De ce point de vue, les sous-graphes isométriques d'hypercubes (aussi appelés *cubes partiels*) peuvent donc être considérés comme une généralisation des graphes médians. La *dimension* $\dim(G)$ d'un graphe médian G est la plus grande dimension d'un hypercube sous-graphe de G .

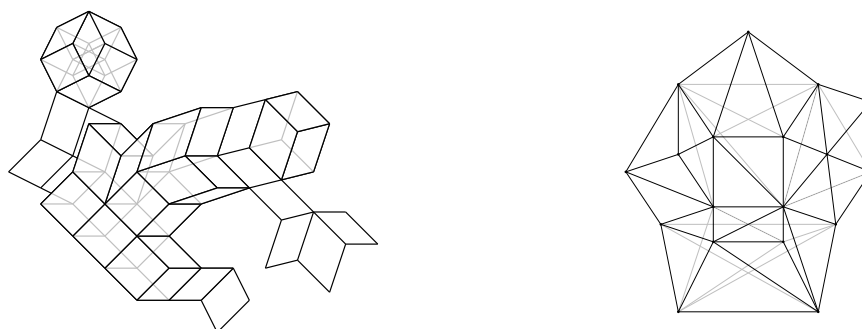


FIGURE 4.2 – Exemple, à gauche, d'un graphe médian quelconque et, à droite, d'un graphe ponté quelconque.

Remarque 10. Le plongement isométrique canonique d'un graphe médian G dans un hypercube (de dimension minimale) peut être déterminé par la relation de parallélisme Θ sur ses arêtes, comme vu en section 1.2.3.

Remarque 11. De même que les retractsés d'hypercubes définissent exactement les graphes médians, les retractsés de graphes de Hamming définissent une autre sous-classe des graphes faiblement modulaires par un résultat de WILKEIT [150], celle des graphes *quasi-médians*, généralisation des graphes médians dans laquelle l'enveloppe convexe de tout C_6 isométrique est un Q_3 , chaque triplet de sommet admet un unique quasi-médian, et dans laquelle aucun $K_{1,1,2}$ n'est induit (c.-à-d., une arête ne peut pas être dans plus d'une clique).

La principale différence entre les graphes médians et les sous-graphes isométriques d'hypercubes généraux, réside dans le fait que les ensembles convexes soient exactement les ensembles portés dans le cas de ces premiers. Par conséquent, tous les demi-espaces $\{W(x, y), W(y, x)\}$ d'un graphe médian G sont portés et, de fait, satisfont la propriété de Helly :

Pour tout sous-ensemble \mathcal{W} de demi-espaces de G , si $\bigcap_{W \in \mathcal{W}} W = \emptyset$, alors il existe $W, W' \in \mathcal{W}$ tels que $W \cap W' = \emptyset$.

En fait, la propriété de Helly des demi-espaces caractérise les graphes médians par un résultat de MULDER et SCHRIJVER [119].

Les graphes médians sont aussi équivalents à des structures apparaissant dans des domaines variés. Pour un aperçu des résultats sur les bijections entre les graphes médians et les algèbres médianes, les demi-treillis médians et les espaces de solutions de formules 2-SAT, voir les papiers de BANDELT et CHEPOI [23], et KNUTH [109]. Pour une présentation des graphes médians en tant que domaines de structures d'évènements (modèle très étudié en théorie de la concurrence [153]), le lecteur pourra aussi consulter la version longue de « A Counterexample to Thiagarajan's Conjecture on Regular Event Structures » [42].

4.2.1 Graphes médians et complexes cubiques CAT(0)

Étant données les propriétés métriques dont disposent les graphes médians et leur relation avec les hypercubes, il paraît naturel d'associer à un graphe médian G un complexe cubique $X(G)$ et d'espérer que ce dernier ait aussi des "bonnes" propriétés structurelles. Concrètement, $X(G)$ est obtenu en remplaçant chaque hypercube sous-graphe de G par un cube solide de la même dimension. G correspond alors au 1-squelette $X^{(1)}(G)$ de $X(G)$ (c.-à-d., au graphe dont les sommets sont les cubes 0-dimensionnels de $X(G)$ et dont les arêtes sont ses cubes 1-dimensionnels). Le complexe cubique $X(G)$ peut être muni de plusieurs métriques intéressantes. Par exemple, la métrique ℓ_1 sur $X(G)$ correspond à la métrique de graphe standard d_G sur G . Une autre métrique importante dans les complexes cubiques X est la métrique ℓ_2 . La distance entre deux sommets $u, v \in X$ associée à cette métrique correspond à la plus grande borne inférieure sur la longueur ℓ_2 d'un chemin reliant u et v . Dans ce cas, un chemin entre u et

v dans X est une séquence $u =: u_0, u_1, \dots, u_k := v$ telle que chaque points consécutif u_i, u_{i-1} appartiennent à un même cube de X (alors (X, ℓ_2) est un espace métrique géodésique¹).

Les complexes cubiques $CAT(0)$ forment une sous-classe importante des complexes cubiques, étudiés en théorie géométrique des groupes. Considérons un espace métrique géodésique (X, d_X) et un triangle géodésique (c.-à-d., triplets de points (x_1, x_2, x_3) reliés par des segments géodésiques) $\Delta := \Delta(x_1, x_2, x_3)$ dans X . Un *triangle de comparaison* de Δ est un triangle $\Delta' := \Delta(x'_1, x'_2, x'_3)$ du plan euclidien \mathbb{E}^2 tel que $d_{\mathbb{E}^2}(x'_i, x'_j) = d_X(x_i, x_j)$ pour tout $i, j \in \{1, 2, 3\}$. L'espace métrique géodésique (X, d_X) est un *espace $CAT(0)$* si tous ses triangles géodésiques Δ satisfont l'*axiome de comparaison Cartan-Alexandrov-Toponogov* :

Si y est un point du côté délimité par x_1 et x_2 de Δ , et si y' est l'unique point sur le segment $[x'_1, x'_2]$ du triangle de comparaison Δ' tel que $d_{\mathbb{E}^2}(x'_i, y') = d_X(x_i, y)$ pour $i = 1, 2$, alors $d_X(x_3, y) \leq d_{\mathbb{E}^2}(x'_3, y')$.

En termes très informels, les triangles des espaces $CAT(0)$ sont plus fins que ceux de l'espace euclidien (comme illustré sur la figure 4.3). Toutefois, pour les complexes cubiques (ou les complexes cellulaires en général), la propriété $CAT(0)$ peut aussi être définie par la propriété simple et intuitive suivante : *toute paire de points est liée par une unique ℓ_2 -géodésique*. En fait, il existe plusieurs manières naturelles de caractériser les espaces $CAT(0)$, et ces derniers ont de nombreuses propriétés géométriques et topologiques (BRIDSON et HAEFLIGER [40] en font une présentation). GROMOV [93] a donné une élégante caractérisation combinatoire des complexes cubiques $CAT(0)$.

Théorème 12 (GROMOV [93]). *Un complexe cubique X muni de la métrique ℓ_2 est $CAT(0)$ si et seulement s'il est simplement connexe (c.-à-d., sans "trous"), et tout triplet de $(k + 2)$ -cubes de X , dont l'intersection est un k -cube, et dont les intersections deux-à-deux sont des $(k + 1)$ -cubes, est contenu dans un $(k + 3)$ -cube de X .*

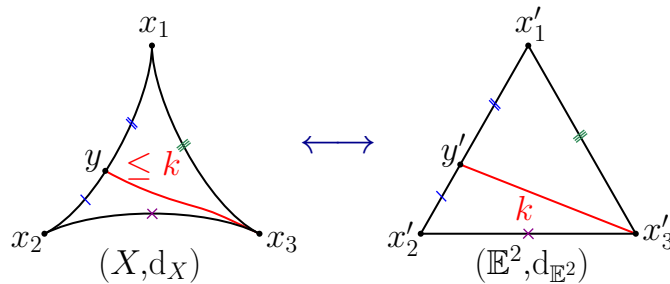


FIGURE 4.3 – Axiome de comparaison Cartan-Alexandrov-Toponogov.

1. Un *segment géodésique* entre deux points u et v de X (muni d'une métrique d_X) est une application $\rho : [a, b] \rightarrow X$ (où $[a, b] \subset \mathbb{R}$, $|a - b| = d_X(u, v)$) telle que, pour tout $s, t \in [a, b]$, $d_X(\rho(s), \rho(t)) = |s - t|$. (X, d_X) est un *espace métrique géodésique* si toute paire de points de X est reliée par un segment géodésique.

Les graphes médians correspondent exactement au 1-squelette des complexes cubiques $CAT(0)$. La preuve de cette correspondance est donnée par CHEPOI [49] via une caractérisation “local-vers-global”, basée sur la caractérisation combinatoire de GROMOV [93] de ces complexes.

Théorème 13 (CHEPOI [49]). *Un graphe G est médian si et seulement si son complexe cubique associé est simplement connexe et satisfait la condition du 3-cube suivante :*

Si trois carrés de G s'intersectent deux à deux sur une arête et ont une intersection globale sur un sommet, alors ces 3 carrés appartiennent à un même 3-cube (voir la figure 4.4 pour une illustration).

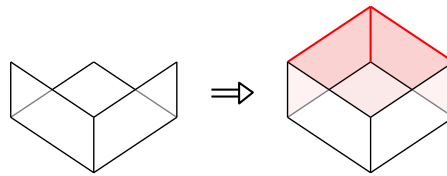


FIGURE 4.4 – Condition du 3-cube.

4.2.2 Problèmes de distances dans les graphes médians et les complexes cubiques $CAT(0)$

La caractérisation de GROMOV [93] a été utilisée pour pointer le fait qu’un grand nombre de complexes cubiques apparaissant en pratique sont $CAT(0)$. BILLERA, HOLMES et VOGTMANN [30] ont démontré que les espaces d’arbres (c.-à-d., sans entrer dans les détails, un complexe cubique dont les faces correspondent aux arbres existants sur un ensemble donné de feuilles) sont des complexes cubiques $CAT(0)$. En fait, les espaces d’arbres sont un cas particulier des bouquets de cubes. ABRAMS et GHRIST [4] et GHRIST et PETERSON [89] ont étudié l’espace continu formé par toutes les positions d’un système reconfigurable (c.-à-d., une collection d’états sur un graphe, distinguant des actions “légalés” selon une base de règles fixée), appelé *espace des états* (“*state complex*”), et ont montré que, dans beaucoup de cas, ce complexe est $CAT(0)$. Motivés par des problèmes de biologie (phylogénie), BILLERA, HOLMES et VOGTMANN [30] ont introduit le problème du calcul des géodésiques (c.-à-d., l’unique plus court chemin) entre deux points d’un espace d’arbres. Dans le domaine de la robotique, les géodésiques dans les espaces d’états $CAT(0)$ correspondent aux “plan de déplacement” d’un point à un autre utilisant un minimum d’énergie. OWEN et PROVAN [127] ont par la suite fourni un algorithme polynomial pour calculer ces géodésiques dans les espaces d’arbres. Dans le cas de complexes de carrés (cubes 2-dimensionnels) $CAT(0)$, le calcul des distances peut se faire en temps linéaire d’après un algorithme de

CHEPOI et MAFTULEAC [60]. Enfin, très récemment, HAYASHI [103] a proposé, pour la première fois, un algorithme polynomial pour le calcul des géodésiques dans les espaces cubiques $CAT(0)$ de dimension quelconque.

En ce qui concerne les graphes médians, le calcul de la distance ou d'un plus court chemin entre deux sommets constitue un problème d'apparence plus simple. Pourtant, pour autant que nous le sachions, aucun algorithme efficace pour faire ce calcul dans le cas général n'est connu.

Remarque 12. Rappelons que CHEPOI, DRAGAN et VAXÈS [53] ont proposé un schéma d'étiquetage de distance utilisant des étiquettes de $O(\log^3 n)$ bits pour les $(3, 6)$ -, $(6, 3)$ - et $(4, 4)$ -graphes planaires (la définition de ces graphes sera donnée plus tard). En particulier, dans le chapitre 7, nous nous intéresserons aux schémas de distance pour les graphes médians sans-cubes, ce sont des $(4, 4)$ -graphes non-planaires.

4.3 Graphes pontés

Les graphes *pontés* ont été introduits par FARBER et JAMISON [73] comme les graphes ne contenant aucun cycle isométrique de plus de trois sommets (un exemple est donné en figure 4.2). Il représentent donc une sur-classe importante de la famille, non moins importante, des graphes *cordaux*. C'est-à-dire, des graphes ne contenant aucun cycle induit de plus de trois sommets. Les graphes cordaux (et donc les graphes pontés) possèdent toujours un sommet *simplicial* (c.-à-d., un sommet dont le voisinage induit une clique), et par conséquent sont démontables par ces sommets.

Comme nous l'avons fait dans la section 4.2, nous commençons par énoncer les propriétés des graphes pontés qui nous fourniront les outils pour les preuves du chapitre 8. Tout d'abord, FARBER et JAMISON [73] et SOLTAN et CHEPOI [143] ont montré que les graphes pontés pouvaient être définis de manière équivalente de la façon qui suit.

Théorème 14. *Un graphe $G = (V, E)$ est ponté si, pour tout sous-ensemble convexe $S \subseteq V$ et pour tout rayon $r \in \mathbb{N}$, la boule $B_r(S, G)$ est convexe.*

Autrement dit, les graphes pontés satisfont l'une des propriétés principales des espaces $CAT(0)$: le voisinage des convexes est convexe. Cette caractérisation a pour conséquence que, dans un graphe ponté $G = (V, E)$, un sous-ensemble $S \subseteq V$ est convexe si et seulement s'il est localement convexe.

Remarque 13. De par la définition des graphes pontés $G = (V, E)$ par la convexité des boules, il suit que :

Si $u, v \in V$ sont deux sommets à distance $k - 1$ d'un sommet central $x \in V$, et s'il sont tous deux voisins d'un sommet $w \in V$ à distance k de x , alors la convexité de $B_{k-1}(x)$ implique que $u \sim v$ (voir la figure 4.5).

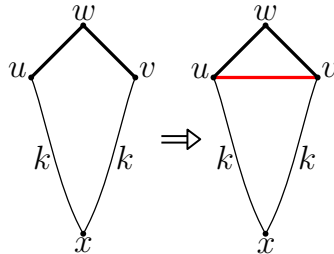


FIGURE 4.5 – Exemple d'une implication de la convexité des boules dans les graphes pontés.

Comme les graphes médians, les graphes pontés jouissent de la propriété suivante :

Lemme 7. *Tout graphe ponté $G = (V, E)$ satisfait les conditions (CT) et (CQ).*

Démonstration. Montrons d'abord que G satisfait la condition du triangle. Soient $u, v, w \in V$ tels que $uv \in E$ et $d_G(u, w) = d_G(v, w) =: k$. Soit x un sommet de $I(u, w) \cap I(v, w)$ à distance minimale de u et v . Alors un cycle C formé par l'arête uv , un (x, u) -chemin et un (x, v) -chemin doit être isométrique par le choix de x . Il s'en suit que C est de longueur 3, et donc que $x \sim u, v$ et que $d_G(w, x) = k - 1$.

Pour montrer que G satisfait aussi la condition du quadrangle. Considérons quatre sommets $u, v, w, z \in V$ tels que $z \sim u, v$, $d_G(u, w) = d_G(v, w) =: k$ et $d_G(w, z) = k + 1$. Alors la convexité de la boule $B_k(w)$ implique que $uv \in E$. Le cas se réduit donc à la condition du triangle, satisfaite par G . \square

Il s'en suit que les graphes pontés constituent une sous-classe des graphes faiblement modulaires. En fait, ils sont caractérisés de la façon suivante :

Théorème 15 (CHEPOI [47]). *Un graphe est ponté si et seulement il est faiblement modulaire et ne contient pas de C_4 ou de C_5 induits.*

Il existe un lien fort entre les graphes pontés et la notion de graphes policier-gagnants introduite par NOWAKOWSKI et WINKLER [125]. Ces graphes sont définis par le jeu du policier et du voleur suivant. Deux joueurs (un policier p et un voleur v) jouent chacun leur tour sur un graphe non-orienté $G = (V, E)$. Au début, p et v sont placés sur des sommets distincts de G et, à chaque tour, ils peuvent soit se déplacer sur un sommet voisin, soit ne pas bouger. Le but de p est

de se retrouver sur le même sommet que v , celui de v est d'empêcher indéfiniment cela. Un graphe G est *policier-gagnant* si, indépendamment du placement initial de p et v , le policier peut toujours gagner. Il se trouve que les graphes pontés sont exactement les graphes policier-gagnants ne contenant pas de C_4 ou de C_5 induits, par un résultat de ANSTEE et FARBER [16]. CHEPOI [48] donne une preuve algorithmique alternative de ce résultat. Cette preuve alternative établit qu'un parcours en largeur sur un graphe ponté $G = (V, E)$ produit systématiquement un "ordonnancement policier-gagnant" de ses sommets (c.-à-d., un ordre total $(V, <)$ dans lequel pour tout $v \in V$, il existe un $v' < v$ adjacent à v tel que pour tout $v'' < v$ adjacent à v , $v'' \sim v'$).

4.3.1 Graphes pontés et complexes systoliques

De même qu'il semble naturel d'associer un complexe cubique à un graphe médian, nous pouvons dériver un complexe simplicial $X(G)$ à partir d'un graphe ponté G en lui associant son *complexe de cliques* : chaque clique de taille k de G est un k -simplexe de $X(G)$, les faces communes de ces simplexes étant déterminées par les intersections des cliques de G . CHEPOI [49] nomme les complexes simpliciaux X dont les graphes pontés sont le 1-squelette des *complexes pontés* ("*bridged complexes*"), et il démontre alors que ces derniers sont simplement connexes et que les "*links*" (c.-à-d., le voisinage ouvert) de leur sommets sont "*flag*" (c.-à-d., X est un complexe de cliques) et sans cycle induit de longueur 4 ou 5. Par la suite, HAGLUND [94] et JANUSZKIEWICZ et ŚWIĄTKOWSKI [104] redécouvrent indépendamment ces complexes simpliciaux dont les graphes pontés sont le 1-squelette. Ils les renomment alors des *complexes (simpliciaux) systoliques*, et c'est plutôt ce terme-ci qui est désormais adopté. JANUSZKIEWICZ et ŚWIĄTKOWSKI [104] définissent les complexes systoliques comme un cas particulier (mais néanmoins important) des complexes *k-systoliques* : complexes simpliciaux simplement connexes et dans lesquels tout cycle de longueur inférieur (strictement) à k dans le "*link*" d'un sommet doit contenir deux arêtes consécutives dans un même 2-simplexe. Les complexes 6-systoliques sont alors nommés *complexes systoliques*. Ce lien entre graphes pontés et complexes systoliques a ensuite donné lieu à une littérature fournie en théorie géométrique des groupes [61, 104, 94, 126].

La proximité des propriétés des complexes systoliques avec celles des complexes cubique CAT(0) fait qu'ils sont considérés comme un analogue aux espaces de courbure non-positive. Ils sont, en particulier, contractiles (voir « Simplicial nonpositive curvature » [104]). Un complexe systolique n'est pas nécessairement hyperbolique en général, mais le seul contre-exemple à cela est le complexe systolique (2-dimensionnel) formé par une triangulation du plan euclidien par des triangles équilatéraux (voir, par exemple, les articles de CHEPOI [49] ou PRZYTY-

CKI [134]). Autrement dit, le complexe systolique associé à un graphe ponté G n'est, en général, pas CAT(0). Il le devient lorsque G est sans K_4 .

Remarque 14. De nouveau, les (3,6)-graphes planaires étudiés par CHEPOI, DRAGAN et VAXÈS [53] pour un schéma de distance contiennent les graphes pontés sans K_4 planaires. Dans le chapitre 8 nous étudions des graphes pontés sans K_4 , ce sont des (3,6)-graphes non-planaires.

4.4 Courbure et théorème de Gauss-Bonnet

Nous rappelons ici un peu de vocabulaire et notations de géométrie (discrète), de sorte à pouvoir énoncer le théorème de Gauss-Bonnet qui nous sera utile dans le chapitre 8. Soit $G = (V, E)$ un graphe planaire. Nous dénotons par ∂G le cycle délimitant sa face externe. Considérons les cycles minimaux (c.-à-d., dont l'intérieur ne contient aucun sommet), de longueur k , de G comme des k -gones réguliers que nous appelons des *faces*. Si nous voyons ces polygones comme des polygones réguliers du plan euclidien, alors la valeur de chaque angle doit être $\frac{k-2}{k}\pi$. Pour tout $v \in G$, $\alpha(v)$ dénote la somme des angles des coins des polygones de G contenant le sommet v . C'est-à-dire que, si l'on pose $\mathcal{C}(v)$ l'ensemble des cycles de G contenant v , alors

$$\alpha(v) := \sum_{C \in \mathcal{C}(v)} \frac{|V(C)| - 2}{|V(C)|} \pi$$

Pour tout $v \in \partial G$, $\tau(v) := \pi - \alpha(v)$, et pour tout $v \in G \setminus \partial G$, $\kappa(v) := 2\pi - \alpha(v)$. Ces valeurs de $\kappa(v)$ ou $\tau(v)$ mesurent donc l'écart entre les angles des polygones autour de v (tels qu'ils sont) et 2π (qui est la valeur qu'ils devraient atteindre si ces polygones étaient "réellement" plongeables dans le plan euclidien). Un *coin* de G est un sommet $v \in \partial G$ tel que $\tau(v) > 0$ et un *sommet interne* de G est un sommet dans $G \setminus \partial G$. Le graphe G sera respectivement dit de *courbure positive*, *nulle* ou *négative* si $\kappa(v) < 0$, $\kappa(v) = 0$ ou $\kappa(v) > 0$ pour tout sommet interne v de G . Une version discrète du théorème de Gauss-Bonnet (voir *Combinatorial Group Theory* [115]) établit la formule suivante (pour un graphe planaire G) :

Théorème 16 (Gauss-Bonnet).

$$\sum_{v \in \partial G} \tau(v) + \sum_{v \in G \setminus \partial G} \kappa(v) = 2\pi.$$

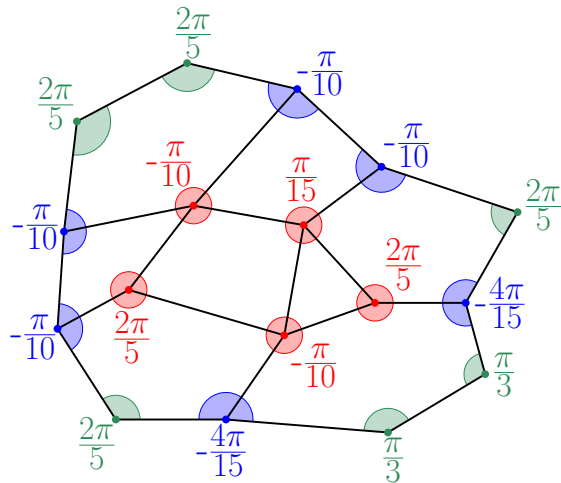


FIGURE 4.6 – Théorème de Gauss-Bonnet et calcul des valeurs de τ et κ pour les sommets d'un graphe planaire G . Les valeurs affichées sont celles de κ pour les sommets internes (en rouge), et celles de τ pour les autres (les coins sont en vert, et les sommets restant en bleu).

$$4\left(-\frac{\pi}{10}\right) + 2\left(-\frac{4\pi}{15}\right) + 4\left(\frac{2\pi}{5}\right) + 2\left(\frac{\pi}{3}\right) + 2\left(-\frac{\pi}{10}\right) + 2\left(\frac{2\pi}{5}\right) + \frac{\pi}{15} = 2\pi.$$

Deuxième partie
Densité et arboricité

Introduction

Dans cette partie, nous nous intéressons à majorer la densité des graphes de deux familles pouvant être vues comme des généralisations des sous-graphes d'hypercubes : les sous-graphes de produits cartésiens de graphes ; et les sous-graphes de demi-cubes. Nous pouvons noter que la motivation initiale des travaux de cette partie est donnée par le théorème 9 de NASH-WILLIAMS [121] et la proposition 6 de KANNAN, NAOR et RUDICH [106] abordés dans le chapitre 3. En effet, ces deux résultats établissent que borner la densité d'un graphe permet directement de déduire une borne supérieure sur la longueur des étiquettes lui étant attribuées par un schéma d'adjacence (basé sur une couverture par des forêts).

Dans le chapitre 5, nous commençons par généraliser les théorèmes 1 et 2 (bornant la densité des sous-graphes d'hypercubes et de graphes de Hamming) aux sous-graphes de produits cartésiens de graphes. Nous étendons ensuite la notion de VC-dimension pour les sous-graphes de produits cartésien d'arêtes (c.-à-d., les hypercubes) aux sous-graphes de produits cartésiens de graphes quelconques. Nous utilisons cette nouvelle définition pour obtenir une généralisation du théorème 8 en adaptant la preuve par induction de HAUSSLER, LITTLESTONE et WARMUTH [100] (présentée en section 2.4.1). Plus précisément, nous démontrons les deux théorèmes suivants :

Théorème (17). *Soit $G = (V, E)$ un sous-graphe d'un produit cartésien de m graphes connexes G_1, \dots, G_m . Alors*

$$\frac{|E|}{|V|} \leq \lceil \max\{\deg^*(G_1), \dots, \deg^*(G_m)\} \rceil \cdot \log |V|,$$

où $\deg^*(G_i)$ dénote le degré moyen maximum de G_i (introduit en section 1.1.3).

Théorème (18). *Soient H un graphe quelconque et $G = (V, E)$ un sous-graphe d'un produit cartésien de m graphes connexes G_1, \dots, G_m n'admettant pas H pour mineur. Alors,*

$$\frac{|E|}{|V|} \leq \mu(H) \cdot \text{VC-dim}^*(G) \leq \mu(H) \cdot \log |V|,$$

où $\mu(H)$ est une constante universelle majorant le degré moyen des graphes excluant le mineur H , et où $\text{VC-dim}^*(G)$ est une notion de VC-dimension généralisée aux produits cartésiens de graphes (que nous introduisons dans le chapitre 5).

En fin de chapitre, nous précisons les bornes supérieures des deux théorèmes précédents pour certaines familles de graphes. Nous nous intéressons en particulier aux sous-graphes de produits cartésien de graphes cordaux (englobant les graphes pontés que nous étudierons dans la partie III), ainsi qu’aux produits cartésiens d’octaèdres.

Comme nous l’avons vu dans le chapitre 1, le théorème 7 de GRAHAM et WINKLER [91] indique que tout graphe connexe et fini G admet un unique plongement isométrique dans un produit cartésien de graphes premiers. Ainsi, nos résultats ont une nature plus générale et montrent qu’il suffit de borner la densité des graphes premiers. Pour la plupart des classes fréquemment étudiée en théorie métrique des graphes, les graphes premiers ont une structure particulière (voir, par exemple, le “survey” de BANDELT et CHEPOI [23]). Par exemple, les graphes premiers pour les sous-graphes isométriques d’hypercubes sont des K_2 . De ce fait, la densité des sous-graphes isométriques d’hypercubes (et, plus généralement, des sous-graphes d’hypercubes) est majorée par leur VC-dimension. Rappelons aussi que SHPECTOROV [141] a démontré que les graphes premiers des graphes admettant un plongement avec distorsion constante dans l’hypercube sont exactement les sous-graphes d’octaèdres et les sous-graphes isométriques de demi-cubes (que nous traitons dans le second chapitre de cette partie (le chapitre 6).

Dans le chapitre 6, nous prouvons un analogue du théorème 8 pour les sous-graphes de demi-cubes. Nous adaptons la notion de VC-dimension à cette famille de graphes, et nous la nommons alors clique-VC-dimension. En nous inspirant de la preuve par décalage de HAUSSLER [99] (présentée en section 2.4.2), nous montrons le théorème suivant :

Théorème (19). *Soit $G = (V, E)$ un sous-graphe de demi-cube induit par une famille d’ensemble \mathcal{S} , et soit d la clique-VC-dimension de \mathcal{S} . Alors*

$$\frac{|E|}{|V|} \leq \binom{d}{2}.$$

Chapitre 5

Sous-graphes de produits cartésiens

Sommaire

5.1	Résultats principaux	73
5.2	Préliminaires	74
5.2.1	Sous-produits, extensions et fibres	74
5.2.2	Mineurs et sous-produits mineurs	76
5.3	VC-dimension et VC-densité	76
5.4	Preuve du théorème 17	79
5.5	Preuve du théorème 18	80
5.5.1	Propriétés de la VC-dimension et de la VC-densité	81
5.5.2	Preuve du théorème 18	87
5.6	Classes spéciales de graphes	88
5.6.1	Graphes de dégénérescence bornée	89
5.6.2	Graphes de degré moyen poly-logarithmique	89
5.6.3	Produits de graphes démontables	90
5.6.4	Produits de graphes cordaux	91
5.6.5	Produits d'octaèdres	91
5.7	Application aux schémas d'adjacence	93

Dans ce chapitre, nous étendons les deux résultats classiques, théorèmes 1 et 8, sur la densité des sous-graphes d'hypercubes aux sous-graphes G de produits cartésiens $G_1 \square \dots \square G_m$ de graphes connexes quelconques. Plus précisément, nous montrons que $\frac{|E(G)|}{|V(G)|} \leq \lceil \max\{\deg^*(G_1), \dots, \deg^*(G_m)\} \rceil \log |V(G)|$. Nous introduisons des notions de VC-dimension $VC\text{-dim}(G)$ et de VC-densité $VC\text{-dens}(G)$ pour les produits cartésiens qui généralisent la notion classique de dimension de Vapnik-Chervonenkis pour les sous-graphes d'hypercubes. Nous prouvons que, si G_1, \dots, G_m appartiennent à la classe $\mathcal{G}(H)$ des graphes connexes finis n'admettant pas le graphe H comme mineur alors, pour tout sous-graphe G de $G_1 \square \dots \square G_m$, l'inégalité plus fine $\frac{|E(G)|}{|V(G)|} \leq \mu(H) VC\text{-dim}(G)$ est vérifiée, où $\mu(H)$ est une constante majorant la densité des graphes de $\mathcal{G}(H)$. Nous raffinons ensuite ces résultats pour des classes de graphes plus spécifiques. Enfin, nous dédui-

sons de nos résultats des bornes (certaines poly-logarithmiques) sur la taille d'étiquettes de schémas d'adjacence pour les sous-graphes de produits cartésiens.

5.1 Résultats principaux

L'objectif principal de ce chapitre est de généraliser les deux théorèmes classiques 1 et 8 sur la densité des sous-graphes d'hypercubes aux sous-graphes G de produits cartésiens $G_1 \square \cdots \square G_m$ de graphes connexes quelconques. Nous allons montrer les deux résultats suivants :

Théorème 17. *Soit G un sous-graphe d'un produit cartésien de m graphes connexes G_1, \dots, G_m . Posons*

$$\begin{aligned} \beta &:= \lceil \max\{\deg^*(G_1), \dots, \deg^*(G_m)\} \rceil \\ \text{et } \beta_0 &:= \lceil \max\{\deg^*(\pi_1(G)), \dots, \deg^*(\pi_m(G))\} \rceil. \end{aligned}$$

Alors

$$\frac{|E(G)|}{|V(G)|} \leq \beta_0 \log |V(G)| \leq \beta \log |V(G)|.$$

Étant donné un graphe H , nous dénotons par $\mathcal{G}(H)$ l'ensemble des graphes finis n'admettant pas de sous-graphe mineur isomorphe à H . Si les facteurs d'un produit cartésien Γ appartiennent tous à $\mathcal{G}(H)$, la généralisation suivante du théorème 8 est vérifiée :

Théorème 18. *Soient H un graphe quelconque et G un sous-graphe d'un produit cartésien Γ de m graphes connexes G_1, \dots, G_m appartenant à $\mathcal{G}(H)$. Alors,*

$$\frac{|E(G)|}{|V(G)|} \leq \mu(H) \cdot \text{VC-dim}^*(G) \leq \mu(H) \cdot \log |V(G)|,$$

où $\mu(H)$ est une constante majorant le degré moyen des graphes de $\mathcal{G}(H)$.

Nous faisons en fait la conjecture plus forte qui suit. Elle sera en partie motivée par le lemme 10 de la section 5.5.1 qui précise que $\text{VC-dens}^*(G) \leq \frac{\mu(H)}{2} \cdot \text{VC-dim}^*(G)$.

Conjecture 1. *Si G est un sous-graphe d'un produit cartésien $G_1 \square \cdots \square G_m$, alors*

$$\frac{|E(G)|}{|V(G)|} \leq \text{VC-dens}^*(G).$$

Dans la section 5.2, nous définissons les sous-produits qui peuvent être vus comme une généralisation des restrictions de coordonnées dans les hypercubes. Comme dans le cas des graphes de 1-inclusion de la section 2.2, nous définissons

ensuite ce que nous appelons des fibres et des extensions (relativement à ces sous-produits). Dans la section 5.3, nous définissons la notions de VC-dimension que nous considérerons. Une preuve du théorème 17 est donnée dans la section 5.4. La preuve du théorème 18 fait l'objet de la section 5.5 dans laquelle nous donnons aussi quelques propriétés basiques des notions de VC-dimension et de VC-densité pour les sous-graphes de produits cartésiens. En section 5.6, nous raffinons les résultats des théorèmes 17 et 18 dans le cas de quelques familles particulières de graphes. Enfin, en section 5.7, nous présentons une application de nos résultats à l'élaboration de borne supérieures sur les étiquettes attribuées par des schémas d'adjacence pour les sous-graphes de produits cartésiens de graphes, ce qui constituait l'une des motivations initiales de ce travail.

5.2 Préliminaires

Nous introduisons maintenant quelques notions préliminaires utiles uniquement à ce chapitre.

5.2.1 Sous-produits, extensions et fibres

Un *sous-produit* Γ' du produit cartésien $\Gamma := \prod_{i=1}^m G_i$ est un produit $\prod_{j=1}^k G'_{i_j}$ dans lequel chaque G'_{i_j} est un sous-graphe connexe non-trivial de G_{i_j} . Un sous-produit $\Gamma' := \prod_{j=1}^k G'_{i_j}$ dans lequel chaque facteur G'_{i_j} est une arête de G_{i_j} est appelé un *sous-cube* de Γ .

Étant donné un sommet $v' = (v'_{i_1}, \dots, v'_{i_k})$ de Γ' , nous dirons que le sommet $v = (v_1, \dots, v_m)$ de Γ est une *extension* de v' si $v_{i_j} = v'_{i_j}$ pour tout $j \in \{1, \dots, k\}$. Nous dénoterons par $F(v')$ l'ensemble de toutes les extensions v dans Γ d'un sommet v' dans un sous-produit $\prod_{j=1}^k G'_{i_j}$ et nous appellerons $F(v')$ la *fibre* de v' dans Γ (une illustration est donnée en figure 5.1).

Soit G un sous-graphe d'un produit cartésien Γ et soit Γ' un sous-produit de Γ . La *trace* de $V(G)$ sur $V(\Gamma')$ consiste en l'ensemble des sommets v' de Γ' ayant une fibre non-vide, (c.-à-d., $F(v') \cap V(G) \neq \emptyset$). La *projection* de G sur Γ' est le sous-graphe $\pi_{\Gamma'}(G)$ de Γ' induit par la trace de $V(G)$ sur $V(\Gamma')$. Pour être plus concis, nous dénoterons la projection de G sur le i -ième facteur G_i de Γ par $\pi_i(G)$ (au lieu de $\pi_{G_i}(G)$).

Le résultat suivant indique que l'ensemble des sous-graphes les plus denses d'un produit cartésien contient des graphes avec une structure de produit cartésien. Plus exactement : la densité d'un produit cartésien de graphes est égale à la densité du produit cartésien des sous-graphes les plus denses de chacun de ses

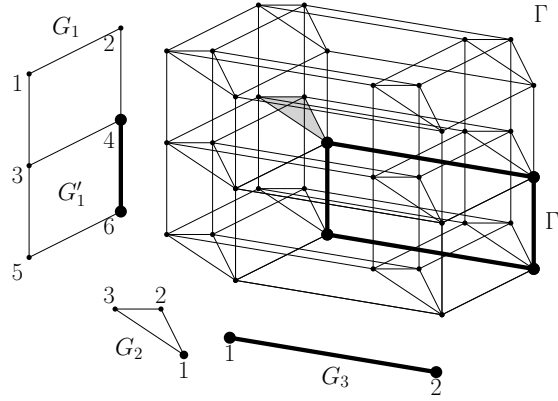


FIGURE 5.1 – Exemple d'un sous-produit (étant aussi un sous-cube) $\Gamma' := G'_1 \square G'_3$ de Γ ainsi que de la fibre (en grise) $F((4,1)) := \{(4,1,1), (4,2,1), (4,3,1)\}$ du sommet $v = (4,1)$ de Γ' . Notons qu'ici $G_3 = G'_3$.

facteurs.

Lemme 8. Soit $\Gamma := \prod_{i=1}^m G_i$, alors $\text{dens}(\Gamma) = \sum_{i=1}^m \text{dens}(G_i)$.

Démonstration. Comme vu dans la proposition 3, le graphe Γ possède $|V_1| \cdots |V_m|$ sommets et $\sum_{i=1}^m |E_i| \cdot (\prod_{j=1, j \neq i}^m |V_j|)$ arêtes, donc $\frac{|E|}{|V|} = \sum_{i=1}^m \frac{|E_i|}{|V_i|} = \sum_{i=1}^m \text{dens}(G_i)$ et, par conséquent, $\text{dens}(\Gamma) \geq \sum_{i=1}^m \text{dens}(G_i)$.

Il nous reste à prouver l'inégalité inverse $\text{dens}(\Gamma) \leq \sum_{i=1}^m \text{dens}(G_i)$. Du fait de l'associativité de l'opération de produit cartésien, il suffit de montrer le résultat pour un produit de deux graphes. Soit $\Gamma = G_1 \square G_2$ et soit $G = (V, E)$ un sous-graphe de Γ parmi les plus denses. Nous qualifierons une arête de G issue de G_1 d'*horizontale* et une arête issue de G_2 de *verticale* et nous noterons E_h et E_v les ensembles respectifs contenant ces arêtes. Nous avons alors $\frac{|E|}{|V|} = \frac{|E_h|}{|V|} + \frac{|E_v|}{|V|}$ et il suffit donc de montrer les inégalités $\frac{|E_h|}{|V|} \leq \text{dens}(G_1)$ et $\frac{|E_v|}{|V|} \leq \text{dens}(G_2)$. Établissons pour commencer la première inégalité. Pour tout sommet $x \in V_2$, nous dénoterons par $L(x) = (V(x), E(x))$ le sous-graphe de G induit par les sommets de G ayant x comme seconde coordonnée. Ce sous-graphe sera nommé *couche* x de G . Toutes les arêtes horizontales et tous les sommets de G sont répartis dans de telles couches. Chaque couche $L(x)$ est un sous-graphe de G_1 , et donc $\text{dens}(L(x)) \leq \text{dens}(G_1)$ pour tout $x \in V_2$. Par conséquent,

$$\frac{|E_h|}{|V|} = \frac{\sum_{x \in V_2} |E(x)|}{\sum_{x \in V_2} |V(x)|} \leq \max_{x \in V_2} \left\{ \frac{|E(x)|}{|V(x)|} \right\} \leq \text{dens}(G_1),$$

comme affirmé¹. La seconde inégalité se montre de la même façon. \square

Autrement dit, le lemme 8 établit que, si les $G'_i = (V'_i, E'_i)$ dénotent des sous-graphes des $G_i = (V_i, E_i)$ de densité maximale, alors $\Gamma' = \prod_{i=1}^m G'_i$ est un sous-graphes de Γ de densité maximale.

5.2.2 Mineurs et sous-produits mineurs

Un *mineur* d'un graphe G est un graphe M pouvant être obtenu comme un sous-graphes G' de G par une suite de contraction d'arêtes. De façon équivalente, M est un mineur d'un graphe connexe G s'il existe une partition des sommets de G en sous-graphes connexes $\mathcal{P} = \{P_1, \dots, P_t\}$ et une bijection $f : V(M) \rightarrow \mathcal{P}$ tels que, si $uv \in E(M)$, alors il existe une arête de G entre les sous-graphes $f(u)$ et $f(v)$ de \mathcal{P} (c.-à-d., après avoir contracté chacun des sous-graphes $P_i \in \mathcal{P}$ en un unique sommet, nous obtenons un graphe couvert par M). Une classe de graphes \mathcal{G} est dite *close par mineurs* si pour tout $G \in \mathcal{G}$ et pour tout mineur M de G , $M \in \mathcal{G}$. Nous pouvons rappeler ici un résultat classique de ROBERTSON et SEYMOUR [136] établissant que *toute famille de graphes close par mineur peut être caractérisée par un ensemble fini de mineurs exclus*.

Par un résultat de Mader, Kostochka et Thomason (voir *Graph Theory* [68, Chapter 8.2]), tout graphe G de degré moyen $\deg(G) \geq cr\sqrt{\log r}$ doit avoir un mineur isomorphe au graphe complet K_r à r sommets. De ce fait, pour tout graphe H avec r sommets, il existe une constante $\mu(H) \leq cr\sqrt{\log r}$ telle que tout graphe $G \in \mathcal{G}(H)$ ait un degré moyen $\deg(G)$ inférieur ou égal à $\mu(H)$.

Un *sous-produit mineur* d'un produit cartésien de graphes connexes $\Gamma := \prod_{i=1}^m G_i$ est un produit cartésien $M := \prod_{i=1}^m M_i$, où M_i est un mineur de G_i pour tout $1 \leq i \leq m$. Posons $\mathcal{P}_i = \{P_1^i, \dots, P_{t_i}^i\}$ la partition de G_i définissant le mineur M_i et $\mathcal{P} := \mathcal{P}_1 \times \dots \times \mathcal{P}_m$. Nous remarquons qu'alors \mathcal{P} est une partition des sommets du produit cartésien $\prod_{i=1}^m G_i$.

5.3 VC-dimension et VC-densité

Nous présentons maintenant les notions de VC-dimension et de VC-densité pour des sous-graphes de produits cartésiens de graphes connexes. Nous définissons dans un premier temps ces notions relativement aux sous-produits. Un sous-produit $\Gamma' := \prod_{j=1}^k G'_{i_j}$ d'un produit cartésien $\Gamma = \prod_{i=1}^m G_i$ est *pulvérisé* par

1. Rappelons que $\sum_{i=1}^m \frac{a_i}{b_i} \leq \max \left\{ \frac{a_i}{b_i} \right\}$ est vérifié pour tout $a_1, \dots, a_m, b_1, \dots, b_m \in \mathbb{N}^*$.

un sous-graphe G de Γ si, pour tout sommet $v' \in V(\Gamma')$, $F(v') \cap V(G) \neq \emptyset$ (c.-à-d., si $\pi_{\Gamma'}(G) = \Gamma'$). La *VC-dimension (induite)* de G respectivement au produit cartésien Γ , notée $\text{VC-dim}(G)$, est le nombre maximum de facteurs non-triviaux (c.-à-d., non réduits à un unique sommet) d'un sous-produit Γ' pris sur tous les sous-produits Γ' de Γ pulvérisés par G . Étant donné que chaque facteur de Γ' contient forcément une arête, nous pouvons définir $\text{VC-dim}(G)$ de façon équivalente comme la plus grande dimension d'un sous-cube de Γ pulvérisé par G . La *VC-densité* $\text{VC-dens}(G)$ de G correspondra à la plus grande densité d'un sous-produit Γ' de Γ pulvérisé par G .

Selon la même idée, nous considérons aussi les notions de VC-dimension et de VC-densité des sous-graphes G d'un produit cartésien Γ respectivement aux sous-produits mineurs de Γ . Soit $M := \prod_{i=1}^m M_i$ un sous-produit mineur de $\Gamma := \prod_{i=1}^m G_i$. Soit $\mathcal{P} := \mathcal{P}_1 \times \cdots \times \mathcal{P}_m$ une partition de $V(\Gamma)$ telle que, pour tout $i \in \{1, \dots, m\}$, $\mathcal{P}_i = \{P_{t_1}^i, \dots, P_{t_i}^i\}$ désigne la partition de G_i définissant le mineur M_i . Nous dirons qu'un sous-produit mineur M est *pulvérisé* par G si tout ensemble $P_{l_1}^1 \times \cdots \times P_{l_m}^m$ de \mathcal{P} (avec $l_1 \in \{1, \dots, t_1\}, \dots, l_m \in \{1, \dots, t_m\}$) contient au moins un sommet de G . La *VC-dimension (mineure)* de G respectivement à Γ , notée $\text{VC-dim}^*(G)$, est le nombre maximum de facteurs non-triviaux d'un sous-produit mineur M pris sur tous les sous-produits mineurs M de Γ pulvérisés par G . De façon analogue, la *VC-densité (mineure)* $\text{VC-dens}^*(G)$ de G est la plus grande densité $\text{dens}(M)$ d'un sous-produit mineur M de Γ pulvérisé par G . La VC-dimension (mineure) peut être définie de façon équivalente comme la plus grande dimension d'un sous-cube mineur (c.-à-d., obtenu en partitionnant chaque facteur en une unique partie ou en deux parties connexes) pulvérisé. La VC-dimension classique (pour les sous-graphes de produits cartésiens d'arêtes) peut être exprimée de la même façon : chaque facteur K_2 est partitionné soit en une unique partie, soit en deux parties d'un sommet.

Remarque 15. Pour un produit cartésien F^m de m fois un graphe donné F , CESA-BIANCHI et HAUSSLER [41] définissent une notion de “*cube projeté d-dimensionnel*” qui coïncide, dans ce cas, avec notre notion de sous-cube pulvérisé.

Remarque 16. De même que pour le cas de la VC-dimension classique, $\text{VC-dim}(G)$ et $\text{VC-dim}^*(G)$ sont définies respectivement à un plongement de G dans un produit cartésien $\prod_{i=1}^m G_i$ fixé. Nous rappelons donc que, lorsque nous parlons de $\text{VC-dim}(G)$ ou $\text{VC-dim}^*(G)$, nous supposons toujours qu'un plongement de G en tant que sous-graphe induit d'un produit cartésien $G_1 \square \cdots \square G_m$ est donné.

Nous motivons maintenant notre choix des noms “VC-dimension” et “VC-densité” dans ce contexte plus général de sous-graphes de produits cartésiens. Nous souhaitons aussi motiver la façon dont nous avons (re)défini ces concepts. Faisons tout d'abord la remarque suivante

Remarque 17. Lorsque G est un sous-graphe de l'hypercube Q_m , les égalités $\text{VC-dim}(G) = \text{VC-dim}^*(G) = \text{VC-dens}(G) = \text{VC-dens}^*(G) = \text{VC-dim}(\mathcal{S}(G))$ sont

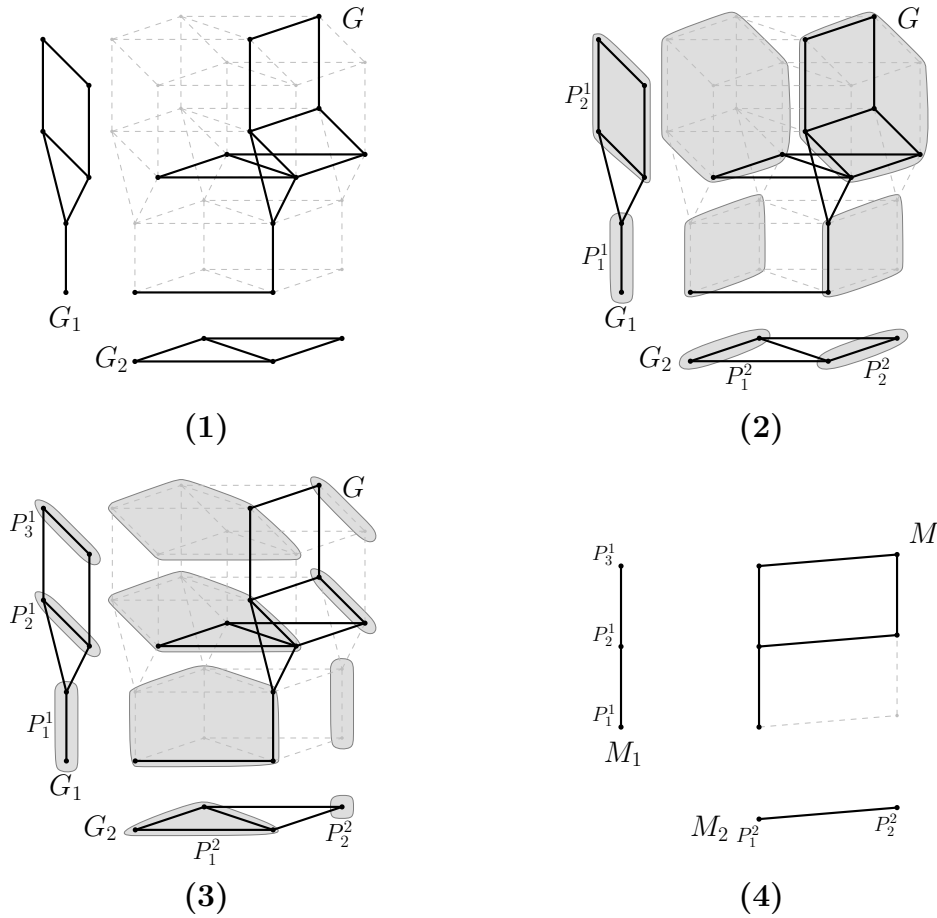


FIGURE 5.2 – Un sous-graphe G de $\Gamma := G_1 \square G_2$ et deux partitions de $V(\Gamma)$, la première induisant un sous-produit mineur pulvérisé par G , la seconde en induisant un non-pulvérisé par G (car $P_1^1 \square P_2^2$ ne contient aucun sommet de G).

vérifiées, et tous ces nombres coïncident avec la dimension du plus grand sous-cube Q de Q_m pulvérisé par G .

Les égalités de la remarque précédente sont vraies car, dans le cas des hypercubes (c.-à-d., produits cartésiens de K_2), les notions de sous-produit, de sous-cube et de sous-produit mineur coïncident. De plus, toutes ces dimensions coïncident avec le degré des sommets de Q , et donc avec le degré moyen $\deg^*(Q)$ de Q . Cela constitue une justification du fait que nous avons défini la VC-densité via le degré moyen d'un sous-produit (mineur ou induit) pulvérisé. D'autre part, le lemme 8 montre que les sous-produits constituent les sous-graphes les plus denses d'un produit cartésien. Aussi, nous pouvons nous attendre à ce que la densité d'un sous-produit le plus dense pulvérisé par un graphe G donne une borne supérieure sur sa densité. Troisièmement, pour les sous-graphes de pro-

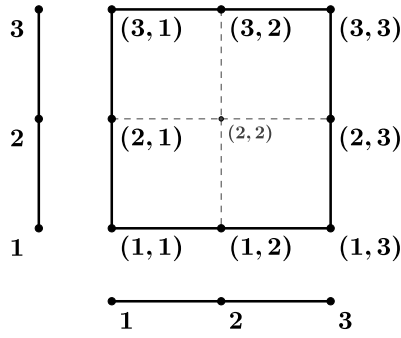


FIGURE 5.3 – Un sous-graphe G de $P_3 \square P_3$ ne pulvérisant aucun sous-produit ayant deux facteurs, mais pulvérisant un sous-produit mineur ayant deux facteurs (induit par $\mathcal{P} = \{\{1, 2\}, \{3\}\} \times \{\{1\}, \{2, 3\}\}$). Autrement dit, $\text{VC-dim}(G) = 1$ et $\text{VC-dim}^*(G) = 2$.

duits cartésiens quelconques, il n'est pas raisonnable de définir $\text{VC-dens}(G)$ ou $\text{VC-dens}^*(G)$ simplement comme le nombre maximal de facteurs non-triviaux dans un sous-produit pulvérisé car ces facteurs pourraient avoir des nombres complètement différents de sommets et d'arêtes. Enfin, nous utilisons le préfixe "VC" car nos dimensions sont définies par une opération de pulvérisation proche de la définition classique.

5.4 Preuve du théorème 17

Posons $n := |V(G)|$. Puisque $\text{dens}(G'_i) \leq \text{dens}(G_i)$ pour tout sous-graphe G'_i de G_i avec $i \in \{1, \dots, m\}$, l'inégalité

$$\lceil \max\{\text{dens}(\pi_1(G)), \dots, \text{dens}(\pi_m(G))\} \rceil \leq \lceil \max\{\text{dens}(G_1), \dots, \text{dens}(G_m)\} \rceil$$

est vérifié. Nous en déduisons la seconde inégalité $\beta_0 \log |V(G)| \leq \beta \log |V(G)|$. Il reste donc à vérifier la première inégalité

$$\frac{|E(G)|}{|V(G)|} \leq \beta_0 \log |V(G)| = \lceil \max\{\text{deg}^*(\pi_1(G)), \dots, \text{deg}^*(\pi_m(G))\} \rceil \cdot \log n$$

par induction sur n . Pour $n = 1$, la propriété est immédiatement vérifiée. Supposons que $n \geq 2$. Alors la projection $\pi_i(G)$ de G sur un certain facteur G_i contient au moins deux sommets d'après le lemme 1. Il en résulte que $\pi_i(G)$ doit avoir deux sommets v'_i et v''_i de degré au plus $\lceil \text{deg}^*(\pi_i(G)) \rceil \leq \beta_0$. Dénotons par A' (resp. A'') l'ensemble des sommets de G ayant v'_i (resp. v''_i) en tant que i -ième coordonnée. Étant donné que $v'_i, v''_i \in V(\pi_i(G))$, ni A' ni A'' ne peuvent être vides. De plus, au moins l'un de ces deux ensembles doit conte-

nir au plus $n/2$ sommets. Disons que $|A'| \leq \frac{n}{2}$. Posons $B := V(G) \setminus A'$ et dénotons par G' et G'' les sous-graphes de G respectivement induits par les ensembles A' et B . Soient $\beta'_0 := \lceil \max\{\deg^*(\pi_1(G')), \dots, \deg^*(\pi_m(G'))\} \rceil$ et $\beta''_0 := \lceil \max\{\deg^*(\pi_1(G'')), \dots, \deg^*(\pi_m(G''))\} \rceil$. Puisque $\pi_j(G')$ et $\pi_j(G'')$ sont des sous-graphes de G_j pour tout $j \in \{1, \dots, m\}$, $\beta'_0, \beta''_0 \leq \beta_0$. Par nos hypothèses d'induction, $|E(G')| \leq \beta'_0 |A'| \log |A'| \leq \beta_0 |A'| \log |A'|$ et $|E(G'')| \leq \beta''_0 |B| \log |B| \leq \beta_0 |B| \log |B|$. Puisque A' et B partitionnent $V(G)$, $E(G)$ est l'union disjointe de $E(G')$, $E(G'')$ et $E(A', B)$, où $E(A', B)$ représente l'ensemble des arêtes de G avec une extrémité dans A' et l'autre dans B . Étant donné que tout sommet de A' possède au plus $\deg(v'_i) \leq \lceil \deg^*(\pi_i(G)) \rceil \leq \beta_0$ arêtes incidentes dans $E(A', B)$, nous obtenons que $|E(A', B)| \leq \beta_0 \cdot |A'| = \beta_0 |A'| \log 2$. Par conséquent, nous avons

$$\begin{aligned} |E(G)| &= |E(G')| + |E(G'')| + |E(A', B)| \\ &\leq \beta_0 |A'| \log |A'| + \beta_0 |B| \log |B| + \beta_0 |A'| \log 2 \\ &= \beta_0 |A'| \log(2|A'|) + \beta_0 |B| \log |B| \\ &\leq \beta_0 |A'| \log\left(2\frac{n}{2}\right) + \beta_0 |B| \log n \\ &= \beta_0 (|A'| + |B|) \log n = \beta_0 n \log n. \end{aligned}$$

□

5.5 Preuve du théorème 18

La preuve du théorème 18 est inspirée de celle du théorème 8 de HAUSSLER, LITTLESTONE et WARMUTH [100] mais est rendue plus technique par la généralisation des concepts. Le schéma de preuve est le suivant. Les trois premiers résultats (lemmes 9, 10 et 11) présentent des propriétés élémentaires de la VC-dimension et de la VC-densité et donnent quelques liens entre ces deux notions. Pour prouver l'analogie dans notre cas du lemme 4, nous étendons les opérations \mathcal{S}_e et \mathcal{S}^e (définies pour des familles d'ensembles dans le chapitre 2, section 2.4.1) aux sous-graphes de produits cartésiens. Pour cela, posons u et v deux sommets adjacents dans un certain facteur G_i , nous allons définir les graphes G_{uv} et G_c^{uv} . Le graphe G_{uv} est obtenu à partir de G en contractant chaque arête de type uv . G_{uv} est donc un sous-graphe du produit cartésien ayant les mêmes facteurs que $\Gamma := G_1 \square \dots \square G_m$ mais dans lequel les arêtes de type uv du i -ième facteur ont été contractées (en supprimant les boucles et arêtes multiples). La définition de G_c^{uv} est un peu plus compliquée et sera donc donnée plus tard. Une fois ces définitions posées, les lemmes 12 et 13 donnent des inégalités analogues à $\text{VC-dim}(\mathcal{S}_e) \leq d$ et $\text{VC-dim}(\mathcal{S}^e) \leq d - 1$ pour les graphes G_{uv} et G_c^{uv} . Enfin, le lemme 14 est une adaptation à G , G_{uv} et G_c^{uv} de l'égalité $|\mathcal{S}| = |\mathcal{S}_e| + |\mathcal{S}^e|$ et de l'inégalité $|E(G)| \leq |E(G_e)| + |E(G^e)| + |V(G^e)|$. Pour montrer par induction

que $\frac{|E(G)|}{|V(G)|} \leq \mu(H) \cdot \text{VC-dim}^*(G)$, nous montrons d'abord dans le lemme 15, établissant que $\frac{|V(G^{uv})|}{|V(G_c^{uv})|} \leq |N| \leq \mu(H) - 1$, où N est l'ensemble des voisins communs à u et v . Ce résultat est vérifié si u ou v a un degré d'au plus $\mu(H)$ (mais un tel sommet doit exister puisque les facteurs de Γ sont $\mu(H)$ -dégénérés).

5.5.1 Propriétés de la VC-dimension et de la VC-densité

Nous poursuivons avec quelques propriétés basiques des VC-dimensions (induites et par mineurs) de produits cartésiens de graphes connexes quelconques et nous étendons le lemme 4. Dans tous nos résultats, nous supposons que G est un sous-graphe avec n sommets d'un produit cartésien $\Gamma := \prod_{i=1}^m G_i = G_1 \square \cdots \square G_m$ de m graphes connexes G_1, \dots, G_m . Puisque la pulvérisation, dans les définitions de $\text{VC-dim}(G)$ et $\text{VC-dens}(G)$, est définie respectivement aux sous-produits, et puisque tous les sous-produits sont des sous-produits mineurs, le résultat suivant est trivial :

Lemme 9. $\text{VC-dim}(G) \leq \text{VC-dim}^*(G)$ et $\text{VC-dens}(G) \leq \text{VC-dens}^*(G)$.

Un exemple d'inégalité stricte entre $\text{VC-dim}(G)$ et $\text{VC-dim}^*(G)$ est donné en figure 5.3.

Remarque 18. La preuve par induction du théorème 18 utilise, entre autre, une opération de contraction (G_{uv}) qui produit des mineurs et se comporte donc assez mal avec la VC-dimension induite (voir la figure 5.5). Dans la section 5.6, nous nous intéressons essentiellement à des familles de graphes dans lesquelles nous pouvons prendre soin que la contraction d'une arête uv donne des sous-graphes (induits), et donc pour lesquelles la VC-dimension induite est adaptée.

Le résultat qui suit justifie en partie que nous formulions la conjecture 1 :

Lemme 10. $\text{VC-dens}^*(G) \leq \frac{\mu(H)}{2} \cdot \text{VC-dim}^*(G)$.

Démonstration. Soient $M := M_{i_1} \square \cdots \square M_{i_k}$ et $M' := M'_{j_1} \square \cdots \square M'_{j_h}$ deux sous-produits mineurs de Γ tels que $\frac{|E(M)|}{|V(M)|} = \text{VC-dens}^*(G)$ et $h = \text{VC-dim}^*(G)$. Premièrement, remarquons ces deux choses : (1) $h \geq k$ car, par définition, nous avons choisi M sans facteurs triviaux et donc si k était plus grand que h , nous aurions au moins choisi $M' = M$; (2) en utilisant la proposition 3 nous obtenons facilement que $\frac{|E(M)|}{|V(M)|} = \sum_{l=1}^k \frac{|E(M_{i_l})|}{|V(M_{i_l})|}$. Puisque nous avons défini Γ comme un produit de graphes sans mineur H , pour tout $l \in \{1, \dots, k\}$, $\frac{|E(M_{i_l})|}{|V(M_{i_l})|} = \frac{1}{2} \deg(M_{i_l}) \leq \frac{\mu(H)}{2}$. Par conséquent,

$$\frac{|E(M)|}{|V(M)|} = \text{VC-dens}^*(G) \leq \frac{\mu(H)}{2} \cdot k \leq \frac{\mu(H)}{2} \cdot h = \frac{\mu(H)}{2} \cdot \text{VC-dim}^*(G).$$

□

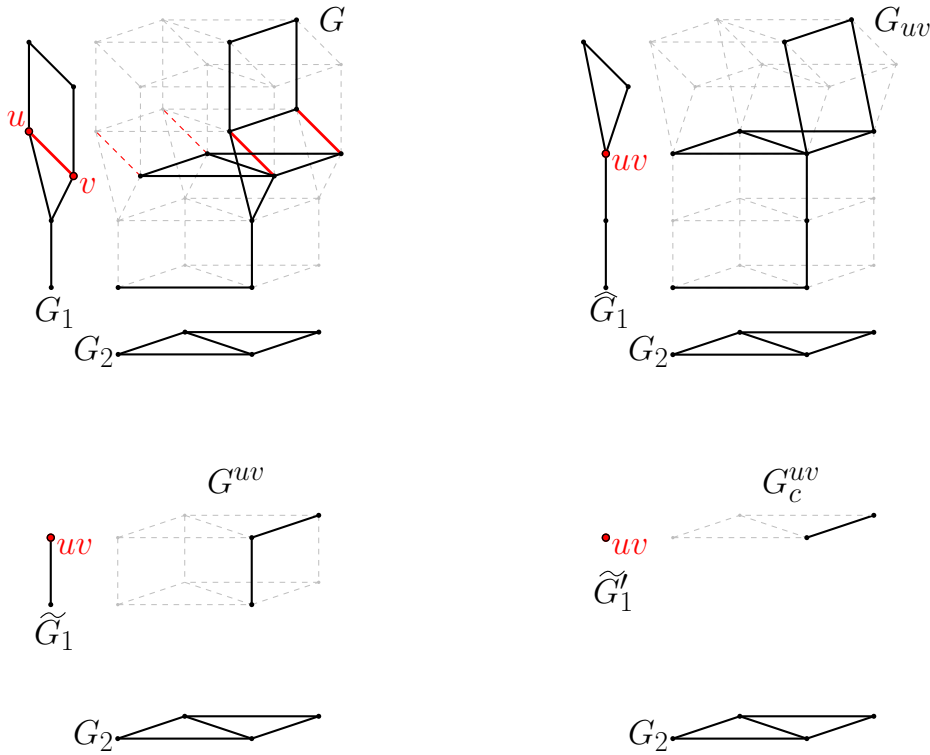
Lemme 11. $\text{VC-dim}^*(G) \leq \log n$.

Démonstration. Soit $M := M_1 \square \cdots \square M_m$ un sous-produit mineur de $\prod_{i=1}^m G_i$ pulvérisé par G avec $k := \text{VC-dim}^*(G)$ facteurs non-triviaux. Nous voulons montrer que $k \leq \log n$. Dénotons par $\mathcal{P}_i = \{P_{t_1}^i, \dots, P_{t_i}^i\}$ la partition de G_i définissant le mineur M_i . Puisque chaque facteur non-trivial de M contient au moins deux sommets, $\mathcal{P} := \mathcal{P}_1 \times \cdots \times \mathcal{P}_m$ contient au moins 2^k parties. Par définition de la pulvérisation, chacune de ces parties $P_{l_1}^1 \times \cdots \times P_{l_m}^m$ (avec $l_1 \in \{1, \dots, t_1\}, \dots, l_m \in \{1, \dots, t_m\}$) contient un sommet x de G . Puisque \mathcal{P} est une partition de Γ , deux sommets appartenant à différentes parties sont nécessairement différents. Nous en déduisons que G contient au moins 2^k sommets. □

Continuons avec l'extension des opérations \mathcal{S}_e et \mathcal{S}^e sur une famille d'ensembles \mathcal{S} aux sous-graphes de produits cartésiens Γ . Dans le cas d'une famille d'ensembles \mathcal{S} , le graphe de 1-inclusion $G(\mathcal{S})$ est un sous-graphe induit d'un produit cartésien d'arêtes (c.-à-d., de K_2). Ainsi, e pointe un facteur complet de ce produit et $G(\mathcal{S}_e)$ peut être vu comme l'image de G dans le produit de K_2 dans lequel un facteur entier (celui pointé par e) a été contracté. Dans le cas où ces facteurs sont des graphes quelconques, contracter un facteur entier serait trop arbitraire. Considérant cela, posons u et v deux sommets adjacents d'un facteur G_i . Soit N l'ensemble des voisins communs à u et v dans G_i . Soit \widehat{G}_i le graphe obtenu depuis G_i en contractant l'arête uv , c'est-à-dire le graphe dans lequel l'arête uv est remplacée par un nouveau sommet w et dans lequel toute arête de la forme xu ou xv de G_i est remplacée par une nouvelle arête xw (sans créer d'arêtes multiples ou de boucles). Soit \widetilde{G}_i l'étoile dont le sommet central est le sommet \widetilde{w} correspondant à l'arête uv et dont l'ensemble \widetilde{N} des sommets feuilles est constitué des sommets \widetilde{x} correspondant aux sommets x de N (c.-à-d., tels que xuv soit un triangle de G_i); les arêtes de \widetilde{G}_i sont les paires de la forme $\widetilde{w}\widetilde{x}$.

Dénotons par G_{uv} le sous-graphe de $\widehat{\Gamma} := G_1 \square \cdots \square G_{i-1} \square \widehat{G}_i \square G_{i+1} \square \cdots \square G_m$ obtenu depuis G par la contraction des arêtes de type uv (c.-à-d., par l'identification des sommets de toutes les paires de la forme $((v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_m), (v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_m))$ et par la suppression des arêtes multiples). Dénotons de même par G^{uv} le sous-graphe de $\widetilde{\Gamma} := G_1 \square \cdots \square G_{i-1} \square \widetilde{G}_i \square G_{i+1} \square \cdots \square G_m$ obtenu depuis G par l'application de la transformation de G_i en \widetilde{G}_i . C'est-à-dire que G^{uv} est le sous-graphe de $\widetilde{\Gamma}$ induit par les règles suivantes :

- (1) $(v_1, \dots, v_{i-1}, \widetilde{w}, v_{i+1}, \dots, v_m)$ est un sommet de G^{uv} si $(v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_m)$ et $(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_m)$ sont des sommets de G ;
- (2) $(v_1, \dots, v_{i-1}, \widetilde{x}, v_{i+1}, \dots, v_m)$ est un sommet de G^{uv} si $(v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_m)$ est un sommet de G avec $x \in N$, et $(v_1, \dots, u, \dots, v_m)$ et $(v_1, \dots, v, \dots, v_m)$ sont aussi des sommets de G .


 FIGURE 5.4 – Exemple des graphes G_{uv} , G^{uv} et G_c^{uv} .

Intuitivement, le graphe G^{uv} correspond, à peu près, au graphe des arêtes perdues par la contraction des arêtes de type uv dans G_{uv} .

Remarquons que G_{uv} joue le rôle de \mathcal{S}_e dans le cas binaire. Pour jouer celui de \mathcal{S}^e , définissons le graphe G_c^{uv} , sous-graphe de G^{uv} induit par les sommets ayant un nœud central \tilde{w} de \tilde{G}_i en tant que i -ième coordonnée. Si G_i est un K_2 (ou, plus généralement, si uv n'appartient à aucun triangle), G^{uv} coïncide avec G_c^{uv} . Les sommets restants de G^{uv} (c.-à-d., ceux ayant une feuille \tilde{x} de \tilde{G}_i en i -ième coordonnée) seront appelés *sommets feuilles*. Nous dénoterons par $V_l(G^{uv})$ l'ensemble des sommets feuilles. L'ensemble des arêtes restantes de G^{uv} , celles de la forme $\tilde{x}\tilde{x}$ ou $\tilde{x}\tilde{w}$, sera noté $E_l(G^{uv})$. Des exemples de graphes G_{uv} et G_c^{uv} sont donnés dans la figure 5.4. Dans tous les résultats qui suivent, nous supposons que G est un sous-graphe d'un produit cartésien $\Gamma := G_1 \square \cdots \square G_m$, que G_i est un facteur quelconque de Γ , et que uv est une arête quelconque de G_i .

Lemme 12. Soit $\Gamma = G_1 \square \cdots \square G_m$ un produit cartésien de graphes finis et soit G un sous-graphe induit de Γ . Pour tout G_i , $i \in \{1, \dots, m\}$, et pour tout $uv \in E(G_i)$,

- (1) $\text{VC-dim}^*(G_{uv}) \leq \text{VC-dim}^*(G)$;
- (2) $\text{VC-dens}^*(G_{uv}) \leq \text{VC-dens}^*(G)$,

où la VC-dimension et la VC-densité de G_{uv} sont calculées par rapport au produit $\hat{\Gamma}$.

Démonstration. Nous voulons montrer que, si un sous-produit mineur M de $\widehat{\Gamma}$ est pulvérisé par G_{uv} , alors M est aussi un sous-produit mineur de Γ pulvérisé par G . Soit $M := M_1 \square \cdots \square M_m$ un sous-produit mineur défini par la partition $\mathcal{P} := \mathcal{P}_1 \times \cdots \times \mathcal{P}_m$ du produit $\widehat{\Gamma}$, où $\mathcal{P}_j := \{P_1^j, \dots, P_{t_j}^j\}$ est une partition de G_j (ou de \widehat{G}_i si $j = i$) définissant le mineur M_j , $j \in \{1, \dots, m\}$. Supposons que M contient k facteurs non-triviaux, indexés par i_1, \dots, i_k . Puisque M est pulvérisé par G_{uv} , tout ensemble $P_{l_1}^1 \times \cdots \times P_{l_m}^m$ de \mathcal{P} (avec $l_1 \in \{1, \dots, t_1\}, \dots, l_m \in \{1, \dots, t_m\}$) contient un sommet x de G_{uv} . Rappelons que nous avons fusionné deux sommets adjacents u et v dans le facteur G_i de Γ pour obtenir G_{uv} . Nous distinguons donc deux cas :

Cas 1. $i \notin \{i_1, \dots, i_k\}$.

Alors M est aussi un sous-produit mineur de Γ . Il nous faut prouver que le sommet x de $P_{l_1}^1 \times \cdots \times P_{l_m}^m$ appartient aussi à G . Considérons pour cela la i -ième coordonnée x_i de x . Si $x_i \neq w$, alors x appartient évidemment à G . Dans le cas contraire, si $x_i = w$, alors au moins un des sommets $(x_1, \dots, u, \dots, x_m)$ ou $(x_1, \dots, v, \dots, x_m)$ doit appartenir à G . Nous pouvons alors considérer que x est l'un de ces sommets.

Cas 2. $i \in \{i_1, \dots, i_k\}$, disons $i := i_j$.

Considérons la partition $\mathcal{P}_i := \{P_1^i, \dots, P_{t_i}^i\}$ de $\widehat{G}_i = \widehat{G}_{i_j}$ et supposons que $w \in P_l^i$ pour un certain $l \in \{1, \dots, t_i\}$. En nous basant sur la partition \mathcal{P}_i de \widehat{G}_i , nous pouvons définir la partition $\mathcal{P}'_i := \{P_1^i, \dots, P_{l-1}^i, P', \dots, P_{t_i}^i\}$ de G_i avec $P' := P_l^i \setminus \{w\} \cup \{u, v\}$. Posons $\mathcal{P}' := \mathcal{P}_1 \times \cdots \times \mathcal{P}'_{i-1} \times \mathcal{P}'_i \times \mathcal{P}_{i+1} \times \cdots \times \mathcal{P}_m$. Il est alors aisé de constater que \mathcal{P}' est une partition du produit cartésien Γ et qu'elle induit une représentation de M dans le produit Γ . De même que dans le cas précédent, il est possible de montrer que l'on a soit $x_i \neq w$ et $x \in V(\Gamma) \cap V(G)$, soit $x_i = w$ et $(x_1, \dots, u, \dots, x_m)$ ou $(x_1, \dots, v, \dots, x_m)$ doit appartenir à G .

Cela montre que, si M est un sous-produit mineur de $\widehat{\Gamma}$ pulvérisé par \widehat{G} , alors M est aussi un sous-produit mineur de Γ pulvérisé par G ; prouvant ainsi les inégalités (1) et (2). \square

Lemme 13. Soit $\Gamma := G_1 \square \cdots \square G_m$ un produit cartésien de graphes finis et soit G un sous-graphe induit de Γ . Pour tout G_i , $i \in \{1, \dots, m\}$, et pour tout $uv \in E(G_i)$,

$$(1) \text{ VC-dim}^*(G_c^{uv}) \leq \text{VC-dim}^*(G) - 1;$$

$$(2) \text{ VC-dens}^*(G_c^{uv}) \leq \text{VC-dens}^*(G) - 1,$$

où la VC-dimension et la VC-densité de G_c^{uv} sont calculées par rapport au produit $\widetilde{\Gamma}$.

Démonstration. Supposons, par l'absurde, que G_c^{uv} pulvérise dans $\widetilde{\Gamma}$ un sous-produit mineur $M := M_1 \square \cdots \square M_m$ ayant soit une densité $\frac{|E(M)|}{|V(M)|}$ au moins égale à $\text{VC-dens}^*(G)$, soit un nombre de facteurs non-triviaux M_{i_1}, \dots, M_{i_k} au moins égal

à $\text{VC-dim}^*(G)$. Dénotons par $\mathcal{P} := \mathcal{P}_1 \times \cdots \times \mathcal{P}_m$ la partition de $\tilde{\Gamma}$ définissant M , où $\mathcal{P}_j := \{P_1^j, \dots, P_{t_j}^j\}$ est une partition de G_j (ou de \tilde{G}_i si $j = i$) définissant le mineur M_j pour chaque $j \in \{1, \dots, m\}$.

Remarquons que tous les sommets de G_c^{uv} doivent avoir \tilde{w} en i -ième coordonnée. De ce fait, si $i \notin \{i_1, \dots, i_k\}$, nous pouvons définir \mathcal{P}' comme le produit de partitions \mathcal{P} dans lequel seule la i -ième partition $\mathcal{P}_i = V(\tilde{G}_i)$ est remplacée par $\mathcal{P}'_i = V(G_i)$. \mathcal{P}' est alors une partition de Γ qui définit le même sous-produit mineur M , maintenant pulvérisé par G . D'autre part, si $i \in \{i_1, \dots, i_k\}$, disons que $i = i_k$, alors étant donné que tous les sommets de G_c^{uv} possèdent \tilde{w} comme i -ième coordonnée, la partition \mathcal{P}_i doit contenir un unique sommet (c.-à-d., $M_i = K_1$ est trivial). Nous pouvons alors supposer que $\mathcal{P}_i = V(\tilde{G}_i)$ et que $i \notin \{i_1, \dots, i_k\}$.

Nous voulons montrer que $M' := M_1 \square \cdots \square M_{i-1} \square K_2 \square M_{i+1} \square \cdots \square M_m$ est un sous-produit mineur de Γ pulvérisé par G . Commençons par préciser la contradiction que cela implique. Puisque le i -ième facteur de M est un K_1 (car $i \notin \{i_1, \dots, i_k\}$), cela implique que $|V(M')| = 2|V(M)|$ et que $|E(M')| = 2|E(M)| + |V(M)|$, conduisant à

$$\begin{aligned} \frac{|E(M')|}{|V(M')|} &= \frac{2|E(M)| + |V(M)|}{2|V(M)|} = \frac{|E(M)|}{|V(M)|} + \frac{1}{2} \\ &= \text{VC-dens}^*(G_c^{uv}) + \frac{1}{2} \\ &\geq \text{VC-dens}^*(G) + \frac{1}{2}, \end{aligned}$$

et menant à une contradiction avec l'hypothèse $\frac{|E(M)|}{|V(M)|} \geq \text{VC-dens}^*(G)$. Cela démontre par la même occasion l'inégalité (2). De plus, $M' = M \square K_2$ impliquerait que $\text{VC-dim}^*(M') = \text{VC-dim}^*(M) + 1$, contredisant l'hypothèse $k \geq \text{VC-dim}^*(G)$ et démontrant (1).

Montrons maintenant que M' est un sous-produit mineur de Γ . Rappelons que $M = M_1 \square \cdots \square M_m$ est défini par la partition $\mathcal{P} := \mathcal{P}_1 \times \cdots \times \mathcal{P}_m$ de $\tilde{\Gamma}$, dans laquelle $\mathcal{P}_i = V(\tilde{G}_i)$. Définissons maintenant la partition $\mathcal{P}' := \mathcal{P}_1 \times \cdots \times \mathcal{P}_{i-1} \times \mathcal{P}'_i \times \mathcal{P}_{i+1} \times \cdots \times \mathcal{P}_m$ de Γ , où $\mathcal{P}'_i = \{P_1^i, P_2^i\}$ et $\{P_1^i, P_2^i\}$ définissent une partition de $V(G_i)$ en deux sous-graphes connexes, le premier contenant u et le second contenant v . Ces parties peuvent être construites de la façon suivante. Posons P_1^i l'ensemble des sommets de G_i pouvant être atteints par un chemin partant de u et ne passant pas par v , et posons $P_2^i := V(G_i) \setminus P_1^i$. M' est alors clairement défini par la partition \mathcal{P}' .

Il reste à montrer que M' est pulvérisé par G dans Γ . Considérons une partie quelconque $P_{l_1}^1 \times \cdots \times P_{l_m}^m$ de \mathcal{P}' (où $l_1 \in \{1, \dots, t_1\}, \dots, l_m \in \{1, \dots, t_m\}$ et $t_i = 2$). M étant pulvérisé par G_c^{uv} , l'ensemble $P_{l_1}^1 \times \cdots \times P_{l_{i-1}}^{i-1} \times V(\tilde{G}_i) \times P_{l_{i+1}}^{i+1} \times \cdots \times P_{l_m}^m$ possède un sommet c appartenant à G_c^{uv} et ayant donc \tilde{w} pour i -ième coordonnée. De par les définitions de G_c^{uv} et du sommet \tilde{w} , nous déduisons que le sommet x_1

obtenu à partir de x en remplaçant la i -ième coordonnée par u est un sommet de $P_{l_1}^1 \times \cdots \times P_1^i \times \cdots \times P_{l_m}^m$. De même, le sommet x_2 obtenu à partir de x en remplaçant la i -ième coordonnée par v appartient à $P_{l_1}^1 \times \cdots \times P_2^i \times \cdots \times P_{l_m}^m$. Puisque x est un sommet de G_c^{uv} , x_1 et x_2 doivent être des sommets de G . Cela montre que les deux parties $P_{l_1}^1 \times \cdots \times P_1^i \times \cdots \times P_{l_m}^m$ et $P_{l_1}^1 \times \cdots \times P_2^i \times \cdots \times P_{l_m}^m$ de \mathcal{P}' contiennent chacune un sommet de G (x_1 et x_2). Par conséquent, G est pulvérisé par M' dans Γ . \square

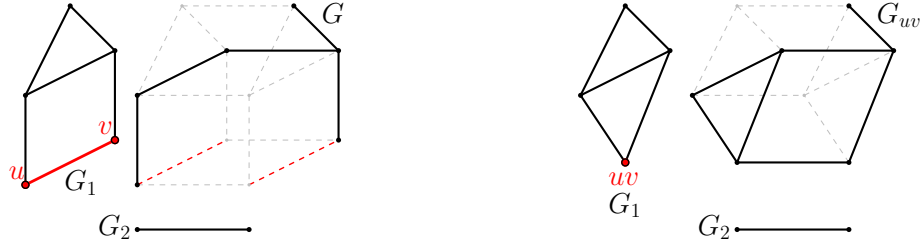


FIGURE 5.5 – Exemple d'un sous-graphe G de $\Gamma := G_1 \square G_2$ et d'un choix d'arête $uv \in E(G_1)$ tels que $\text{VC-dim}(G) < \text{VC-dim}(G_{uv})$. En effet, Γ n'étant constitué que de deux facteurs, pour atteindre une VC-dimension induite de 2, il faudrait que G contiennent tous les sommets d'un produit cartésien d'une arête de G_1 et d'une arête de G_2 . Ce n'est pas le cas, donc $\text{VC-dim}(G) = 1$, mais $\text{VC-dim}(G_{uv}) = 2$.

Lemme 14. *Les graphes G, G_{uv}, G_c^{uv} et G^{uv} satisfont la relation suivante :*

$$\begin{cases} |V(G)| = |V(G_{uv})| + |V(G^{uv})| - |V_l(G^{uv})| = |V(G_{uv})| + |V(G_c^{uv})|, \\ |E(G)| \leq |E(G_{uv})| + |E(G^{uv})| + |V(G_c^{uv})|. \end{cases}$$

Démonstration. G_{uv} contient les sommets de G moins un sommet pour chaque arête de type uv contractée. G^{uv} contient un sommet pour chaque arête contractée de G (les sommets dans $V(G_c^{uv})$) ainsi qu'un sommet pour chaque triangle de G contenant une arête de type uv (les sommets de $V_l(G^{uv})$). Nous obtenons donc que $|V(G)|$ est égal à $|V(G_{uv})| + |V(G^{uv})|$ moins les sommets $|V_l(G^{uv})|$.

$E(G_{uv})$ contient les arêtes de G moins (1) les arêtes de type uv contractées, moins (2) les arêtes multiples obtenues par la contraction des triangles de G contenant une arête de type uv , moins (3) les arêtes multiples obtenues en contractant les carrés de G possédant deux arêtes de type uv opposées, plus (4) d'éventuelles arêtes qui auront pu être créées par la projection d'un sommet v opposé à un sommet u dans un carré de Γ ne possédant que ces deux sommets dans G . Remarquons qu'il y a $|V(G_c^{uv})|$ arêtes de type uv dans le groupe (1), $|E_l(G^{uv})|$ dans le groupe (2) et $|E(G_c^{uv})|$ dans le troisième groupe. Puisque $|E(G^{uv})| = |E_l(G^{uv})| + |E(G_c^{uv})|$, nous obtenons l'inégalité annoncée $|E(G)| \leq |E(G_{uv})| + |E(G^{uv})| + |V(G_c^{uv})|$. \square

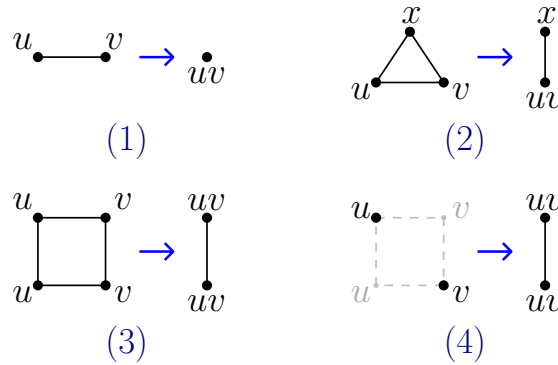


FIGURE 5.6 – Illustration des cas (1) à (4) du dénombrement des arêtes de $E(G_{uv})$ dans le lemme 14.

Lemme 15. $\frac{|V(G^{uv})|}{|V(G_c^{uv})|} \leq |N|$, où N est l'ensemble des voisins communs à u et v .

Démonstration. Chaque sommet de G^{uv} est soit un sommet central de la forme $(v_1, \dots, v_{i-1}, \tilde{w}, v_{i+1}, \dots, v_m)$, soit un sommet feuille de la forme $(v_1, \dots, v_{i-1}, \tilde{x}, v_{i+1}, \dots, v_m)$. Chaque feuille de G^{uv} ne peut être adjacente qu'à un unique sommet central (et ne peut être dans le graphe qu'à condition que le centre en question y soit aussi, par la condition (2) dans la définition de G^{uv}). De plus, chaque sommet central de G^{uv} est adjacent à au plus $|N|$ feuilles, le résultat est donc démontré. \square

5.5.2 Preuve du théorème 18

Soit H un graphe quelconque et soit G un sous-graphe à n sommets d'un produit cartésien $\Gamma := G_1 \square \dots \square G_m$ de m graphes connexes $G_1, \dots, G_m \in \mathcal{G}(H)$. Nous voulons montrer que $\frac{|E(G)|}{|V(G)|} \leq \mu(H) \cdot \text{VC-dim}^*(G) \leq \mu(H) \cdot \log n$. La seconde inégalité découle directement du lemme 11. Il reste à démontrer que $\frac{|E(G)|}{|V(G)|} \leq \mu(H) \cdot \text{VC-dim}^*(G)$ par induction sur les nombres de sommets dans les facteurs de Γ . Puisque chaque facteur G_i de Γ appartient à $\mathcal{G}(H)$, G_i est $\mu(H)$ -dégénéré (c.-à-d., G_i ainsi que tous ses sous-graphes contiennent un sommet v de degré au plus $\mu(H)$). Soit u un voisin de v dans G_i . Alors l'ensemble N des voisins communs à u et v a un cardinal d'au plus $\mu(H) - 1$. Considérons les graphes G_{uv} , G^{uv} et G_c^{uv} obtenus à partir de G en appliquant les opérations définies dans les sections précédentes, respectivement à l'arête uv de G_i . Alors G_{uv} est un sous-graphe du produit $\hat{\Gamma} := G_1 \square \dots \square G_{i-1} \square \hat{G}_i \square G_{i+1} \dots \square G_m$. Vu que \hat{G}_i est un mineur de G_i , tous les facteurs de $\hat{\Gamma}$ appartiennent à $\mathcal{G}(H)$. De plus, puisque \hat{G}_i contient moins de sommets que G_i , nous pouvons appliquer les hypothèses d'induction aux sous-graphes de $\hat{\Gamma}$, en particulier à G_{uv} . De même, G^{uv} et G_c^{uv} étant des sous-graphes de $\tilde{\Gamma} = G_1 \square \dots \square G_{i-1} \square \tilde{G}_i \square G_{i+1} \square \dots \square G_m$, et étant donné que \tilde{G}_i est une étoile isomorphe à un sous-graphe de G_i , tous les facteurs de $\tilde{\Gamma}$ appartiennent

aussi à $\mathcal{G}(H)$. Puisque \tilde{G}_i contient moins de sommets que G_i , les graphes G^{uv} et G_c^{uv} aussi. Par conséquent, les inégalités $\frac{|E(G_{uv})|}{|V(G_{uv})|} \leq \mu(H) \cdot \text{VC-dim}^*(G_{uv})$ et $\frac{|E(G_c^{uv})|}{|V(G_c^{uv})|} \leq \mu(H) \cdot \text{VC-dim}^*(G_c^{uv})$ sont vérifiées.

En utilisant le lemme 14 et l'inégalité $\frac{a_1+a_2}{b_1+b_2} \leq \max\{\frac{a_1}{b_1}, \frac{a_2}{b_2}\}$, nous obtenons

$$\begin{aligned} \frac{|E(G)|}{|V(G)|} &\leq \frac{|E(G_{uv})| + |E(G^{uv})| + |V(G_c^{uv})|}{|V(G_{uv})| + |V(G_c^{uv})|} \\ &\leq \max\left\{ \frac{|E(G_{uv})|}{|V(G_{uv})|}, \frac{|E(G^{uv})| + |V(G_c^{uv})|}{|V(G_c^{uv})|} \right\}. \end{aligned}$$

Par le lemme 12, $\text{VC-dim}^*(G_{uv}) \leq \text{VC-dim}^*(G)$, d'où, par hypothèses d'induction,

$$\frac{|E(G_{uv})|}{|V(G_{uv})|} \leq \mu(H) \cdot \text{VC-dim}^*(G_{uv}) \leq \mu(H) \cdot \text{VC-dim}^*(G).$$

Il reste donc à trouver une borne supérieure similaire pour $\frac{|E(G^{uv})| + |V(G_c^{uv})|}{|V(G_c^{uv})|}$.

Puisque $|E(G^{uv})| + |V(G_c^{uv})| = |E(G_c^{uv})| + |V_i(G^{uv})| + |V(G_c^{uv})| = |E(G_c^{uv})| + |V(G^{uv})|$ et $|N| \leq \mu(H)$, par les lemmes 13 et 15 nous concluons que

$$\begin{aligned} \frac{|E(G^{uv})| + |V(G_c^{uv})|}{|V(G_c^{uv})|} &= \frac{|E(G_c^{uv})| + |V(G^{uv})|}{|V(G_c^{uv})|} \\ &\leq \mu(H) \cdot \text{VC-dim}^*(G_c^{uv}) + \mu(H) \\ &\leq \mu(H) \cdot (\text{VC-dim}^*(G) - 1) + \mu(H) \\ &= \mu(H) \cdot \text{VC-dim}^*(G). \end{aligned}$$

Cela démontre l'inégalité $\frac{|E(G)|}{|V(G)|} \leq \mu(H) \cdot \text{VC-dim}^*(G)$ et termine la preuve du théorème. \square

Le théorème 18 utilisé avec le lemme 3.1 de « Colorings and orientations of graphs » [7] nous donne directement le corollaire suivant :

Corollaire 2. *Si G est un sous-graphe d'un produit cartésien de graphes connexes appartenant à $\mathcal{G}(H)$ et si $d := \text{VC-dim}^*(G)$ dénote sa VC-dimension, alors G peut être orienté de telle sorte que le degré sortant de tout sommet soit majoré par $d\mu(H)$.*

5.6 Classes spéciales de graphes

Dans cette section, nous précisons (et améliorons dans certain cas) nos résultats pour des produits cartésiens de graphes appartenant à des classes plus

spécifiques. Plus précisément, nous étudions les produits cartésiens de graphes de dégénérescence bornée, de degré moyen poly-logarithmique, démontables, cordaux, complets et d'octaèdres.

5.6.1 Graphes de dégénérescence bornée

Nous dénotons ici par \mathcal{G}_δ la classe des graphes de dégénérescence au plus δ . \mathcal{G}_δ contient donc en particulier les graphes dans lesquels chaque sommet a un degré d'au plus δ . Ainsi, \mathcal{G}_5 inclue la classe des graphes planaires. Tout graphe de \mathcal{G}_δ a densité au plus δ . En utilisant le théorème 17, nous obtenons directement le corollaire suivant :

Corollaire 3. *Si G est un sous-graphe de $G_1 \square \cdots \square G_m$ avec $G_1, \dots, G_m \in \mathcal{G}_\delta$, alors $\frac{|E(G)|}{|V(G)|} \leq 2\delta \cdot \log |V(G)|$.*

5.6.2 Graphes de degré moyen poly-logarithmique

Dénotons maintenant par $\mathcal{G}_{(\log, k)}$ la classe des graphes clos par sous-graphes et de \deg^* borné par une fonction poly-logarithmique (c.-à-d., pour tout $H \in \mathcal{G}_{(\log, k)}$, $2 \lceil \text{dens}(H) \rceil \leq \log^k |V(H)|$). Dans ce cas, le théorème 17 conduit au résultat suivant

Corollaire 4. *Si G est un sous-graphe de $\Gamma := G_1 \square \cdots \square G_m$ avec $G_1, \dots, G_m \in \mathcal{G}_{(\log, k)}$, alors $G \in \mathcal{G}_{(\log, k+1)}$. Ou encore, $\frac{|E(G)|}{|V(G)|} \leq \log^{k+1} |V(G)|$.*

Démonstration. D'après le théorème 17,

$$\frac{|E(G)|}{|V(G)|} \leq \max\{\lceil \deg^*(\pi_1(G)) \rceil, \dots, \lceil \deg^*(\pi_m(G)) \rceil\} \cdot \log |V(G)|.$$

Puisque $\pi_i(G)$ est un sous-graphe de G_i , toutes les projections $\pi_i(G)$, $i \in \{1, \dots, m\}$, de G sur les facteurs de Γ appartiennent à $\mathcal{G}_{(\log, k)}$. Nous en déduisons que

$$\lceil \deg^*(\pi_i(G)) \rceil \leq \log^k |V(\pi_i(G))| \leq \log^k |V(G)|,$$

et donc que

$$\frac{|E(G)|}{|V(G)|} \leq \log^{k+1} |V(G)|.$$

□

5.6.3 Produits de graphes démontables

Un sommet u d'un graphe G est dit *dominé* dans G par son voisin v si $N[u] \subseteq N[v]$ (c.-à-d., le voisinage fermé de u est inclus dans celui de v). Dans ce cas, l'arête uv est appelée *arête dominante*. Un graphe G est *démontable* si ses sommets admettent un ordonnancement v_1, \dots, v_n tel que, pour tout i , v_i soit dominé dans le sous-graphe $G[v_i, \dots, v_n]$ de G induit par les sommets v_i, \dots, v_n . La classe des graphes démontables inclut, en particulier, les classes des graphes cordaux, pontés (voir la section 4.3) et faiblement pontés (une définition de ces deux dernières classes ainsi que des résultats sur leur démontabilité sont disponibles dans « Bucolic complexes » [37]). Les graphes démontables sont aussi caractérisés par NOWAKOWSKI et WINKLER [125] comme étant exactement les graphes policier-gagnants.

Un *ordre minimal de démontage* de G est un ordre de démontage dans lequel, à chaque étape i , v_i est un sommet dominé dans $G[v_i, \dots, v_n]$ de degré minimal. Nous définissons la *dégénérescence de démontage* $\text{dd}(G)$ de G comme le degré maximum d'un sommet v_i de $G[v_i, \dots, v_n]$ dans un ordre minimal de démontage de G . Soit $\Gamma := \prod_{i=1}^m G_i$ un produit cartésien de graphes démontables G_1, \dots, G_m . Soit $v_{i,1}, \dots, v_{i,n_i}$ un ordonnancement des sommets de G_i selon un ordre minimal de démontage $<_i$ ($i \in \{1, \dots, m\}$). Soit $(\prod_{i=1}^m V_i, \preceq)$ l'ordre partiel sur les sommets de Γ défini par le produit cartésien des ordres totaux $(V_i, <_i)$ ($i \in \{1, \dots, m\}$). Considérons une extension linéaire quelconque $<$ de \preceq . Le degré maximal d'un sommet v_i d'un sous-graphe de Γ induit par v_i, \dots, v_N , où $N = n_1 \times \dots \times n_m$, sera noté $\text{dd}^*(\Gamma)$. Nous pouvons alors facilement montrer que $\text{dd}^*(\Gamma) = \sum_i \text{dd}(G_i)$.

Pour un sous-graphe G de Γ avec des sommets v_1, \dots, v_n ordonnés selon l'ordre total $<$ sur $V(\Gamma)$, nous définissons $\text{dd}^*(G)$ comme le degré maximal d'un sommet v_i dans le sous-graphe de G induit par v_i, \dots, v_n . Remarquons que $\text{dd}^*(G) \leq \text{dd}^*(\Gamma)$, et que contracter une arête dominante $u_i v_i$ d'un facteur G_i donne un mineur de G_i qui est aussi un sous-graphe induit de G_i . Ainsi, si G est un sous-graphe d'un produit cartésien de graphes démontables, et si à chaque étape de la preuve du théorème 18 nous prenons soin de ne contracter que des arêtes dominées dans leur facteur, nous obtenons le résultat suivant :

Proposition 8. *Si G est un sous-graphe d'un produit cartésien Γ de m graphes démontables G_1, \dots, G_m , alors $\frac{|E(G)|}{|V(G)|} \leq \text{dd}^*(G) \cdot \text{VC-dim}(G)$.*

Démonstration. Puisque G^{uv} est un sous-graphe de G , par le lemme 15, nous déduisons que $\frac{|V(G^{uv})|}{|V(G_c^{uv})|} \leq \text{dd}^*(G^{uv}) \leq \text{dd}^*(G)$. Puisque G_{uv} est un sous-graphe de G , nous obtenons directement que $\text{VC-dim}(G_{uv}) \leq \text{VC-dim}(G)$. Ainsi, en utilisant le lemme 13 pour $\text{VC-dim}(G)$, nous avons $\text{VC-dim}(G_c^{uv}) \leq \text{VC-dim}(G) - 1$. Le résultat suit alors de cette inégalité et des deux remarques qui la précèdent. \square

5.6.4 Produits de graphes cordaux

Pour un graphe cordal G (défini en section 4.3) nous utilisons le fait que tout ordonnancement simplicial des sommets de G est un ordre de démontage, et que pour toute arête uv d'un graphe cordal, G_{uv} est cordal et $\omega(G_{uv}) \leq \omega(G)$ (où $\omega(G)$ est la taille d'une plus grosse clique de G). Comme dans le cas des graphes démontables, en contractant à chaque étape une arête entre un sommet *simplicial* (c.-à-d., dont le voisinages induit une clique) et un de ses voisins, et en appliquant cette propriété de démontabilité, nous obtenons le corollaire suivant de la proposition 8 et du théorème 18 :

Corollaire 5. *Si G est un sous-graphe d'un produit cartésien Γ de m graphes cordaux G_1, \dots, G_m , alors $\frac{|E(G)|}{|V(G)|} \leq \omega(G) \cdot \text{VC-dim}(G)$.*

Un sous-ensemble “classique” de la classe des produits cartésiens de graphes cordaux est la classe des graphes de Hamming. De ce fait, le corollaire 5 s'applique aussi pour la classe des sous-graphes de graphes de Hamming.

5.6.5 Produits d'octaèdres

L'octaèdre d -dimensionnel $K_{2,\dots,2}$ (équivalent au “cocktail party graph” introduit par DEZA et LAURENT [66]) est un graphe biparti complet sur $2d$ sommets moins un couplage parfait. De façon équivalente, $K_{2,\dots,2}$ est un graphe d -parti complet dans lequel chaque partie est constituée de 2 sommets que nous dirons *opposés*. En géométrie, $K_{2,\dots,2}$ correspond aussi au 1-squelette d'un hyperoctaèdre d -dimensionnel. Nous utiliserons la notation $e_2 = \bar{e}_1$ pour indiquer que e_1 et e_2 sont deux sommets opposés. Un sous-graphe d'octaèdre sera appelé *sous-octaèdre*. L'octaèdre d -dimensionnel $K_{2,\dots,2}$ peut aussi être vu comme le graphe d'un alphabet fini $\Sigma = \{e_1, \bar{e}_1, \dots, e_d, \bar{e}_d\}$ muni d'une involution $\varphi(e_i) = \bar{e}_i$ pour tout $i \in \{1, \dots, d\}$ (rappelons qu'une *involution* sur un ensemble fini X est une application $\varphi : X \rightarrow X$ telle que $\varphi^2(x) = x$ pour tout $x \in X$). Par conséquent, le produit cartésien de m octaèdres d -dimensionnels peut être vu comme le graphe des mots de taille m sur (Σ, φ) .

Dans cette sous-section, nous montrons que, si G est un sous-graphe d'un produit cartésien Γ de m octaèdres G_1, \dots, G_m de dimensions respectives d_1, \dots, d_m , alors un analogue du corollaire 5 est vérifié. Pour le problème arête-isopérimétrique sur des produits d'octaèdres, nous pourrions nous référer au livre de HARPER [98]. La principale différence avec les graphes cordaux dans notre cas réside dans le fait que contracter une arête dans un octaèdre peut augmenter la taille de sa plus grande clique. De plus, le résultat de cette contraction n'est plus un sous-octaèdre. De ce fait, nous devons redéfinir les graphes G_{uv} et G^{uv} . Si, au

début, tous les facteurs sont des octaèdres, il est évident qu'après un certain nombre d'étapes, ils deviendront des sous-octaèdres. Lorsqu'un facteur n'est pas une clique, il contient deux sommets opposés e et \bar{e} que nous pouvons identifier, transformant ainsi le facteur en un sous-octaèdre avec moins de paires d'opposés. Si tous les facteurs sont des cliques, ce qui doit finir par arriver, alors leur produit est un graphe de Hamming (qui doit alors être vu comme notre base d'induction) et nous pouvons utiliser le corollaire 5 de la sous-section précédente.

Soit G_i un facteur du produit cartésien courant Γ . Supposons que G_i soit un sous-octaèdre contenant au moins deux sommets opposés e et \bar{e} . Soit F un sous-graphe de Γ . Nous dénoterons par v^i un sommet de Γ dont toutes les composantes sont fixées à l'exception de la i -ième (c.-à-d., v^i fixe la position d'une copie de G_i , mais pas le sommet considéré dans cette copie). Nous dénoterons ce sommet par $v^i[e]$, $v^i[\bar{e}]$, ou $v^i[e']$ si nous voulons respectivement affecter à cette i -ième coordonnée e , \bar{e} ou un certain voisin $e' \in V(G_i)$ commun à e et \bar{e} . Posons \widehat{G}_i le graphe induit par $V(G_i) \setminus \{\bar{e}\}$ et posons G_e le sous-graphe de $\widehat{\Gamma}$ dans lequel nous avons identifié (fusionné) les sommets de G ayant e en i -ième coordonnée avec ceux ayant \bar{e} . Posons maintenant \widetilde{G}_i l'étoile de centre \bar{e} et dont les feuilles sont les voisins de \bar{e} dans G_i . G^e correspond alors au sous-graphe de $\widetilde{\Gamma}$ satisfaisant les deux propriétés suivantes :

- (1) $v^i[\bar{e}] \in V(G^e)$ si et seulement si $v^i[e], v^i[\bar{e}] \in V(G)$;
- (2) $v^i[e'] \in V(G^e)$ si et seulement si $v^i[e'], v^i[\bar{e}], v^i[e] \in V(G)$.

À nouveau, l'idée est d'inclure dans G^e les arêtes de G perdues par l'opération de fusion.

Lemme 16.
$$\begin{cases} |V(G)| = |V(G_e)| + |V(G_c^e)| \\ |E(G)| \leq |E(G_e)| + |E(G_c^e)| + |V_l(G^e)|. \end{cases}$$

Démonstration. Le comptage des sommets de G dans ce cas est identique à celui effectué dans la preuve du lemme 14. Les changements concernent le comptage des arêtes. Le lien entre l'ensemble $E(G)$ et les ensembles $E(G^e)$, $E(G_e)$ et $V_l(G^e)$ sont illustrés dans la figure 5.7. Pour être plus précis, si $v^i[e], v^i[\bar{e}] \in V(G)$, alors les arêtes entre ces sommets et leurs voisins dans les copies indexées par v^i sont comptées une fois dans $E(G_e)$ et une fois dans $V_l(G^e)$, et les arêtes de ces sommets vers les autres copies sont comptées une fois dans $E(G_c^e)$ et une fois dans $E(G_e)$. De façon analogue au cas du lemme 14, une nouvelle arête pourrait être créée si pour un certain couple v_1^i et v_2^i , $v_1^i[e], v_2^i[\bar{e}] \in V(G)$ mais $v_1^i[\bar{e}], v_2^i[e] \notin V(G)$. \square

En reprenant les arguments de la section 5.5.1, nous pouvons montrer que (1) $\text{VC-dim}(G_e) \leq \text{VC-dim}(G)$ et que (2) $\text{VC-dim}(G_c^e) \leq \text{VC-dim}(G) - 1$. Remarquons aussi que $|V(G_c^e)| = 0$ si et seulement si G^e est vide. Autrement, si une feuille $v^i[e']$ existe dans G^e , le nœud central $v^i[\bar{e}]$ doit aussi exister (par la seconde condition de la définition). Ces feuilles sont les voisins (dans les copies de \widetilde{G}_i indexées par

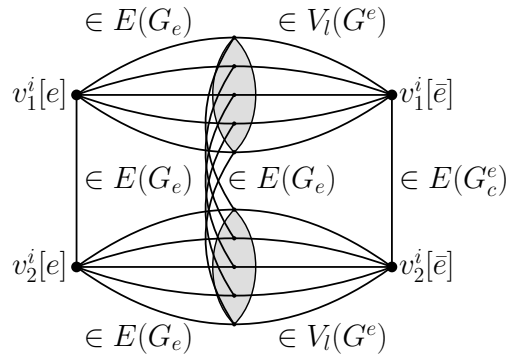


FIGURE 5.7 – Pour la preuve du lemme 16.

v^i) de $v^i[\bar{e}]$ dans G^e et, clairement, elle ne peuvent pas être d'un nombre plus grand que $\omega(G)$. Donc $\frac{|V_l(G^e)|}{|V(G^e)|} \leq \omega(G)$. Finalement, en utilisant le lemme 16, les inégalités (1) et (2) et cette dernière inégalité, nous obtenons

Proposition 9. $\frac{|E(G)|}{|V(G)|} \leq \omega(G) \cdot \text{VC-dim}(G)$.

5.7 Application aux schémas d'adjacence

En utilisant le théorème 9 de NASH-WILLIAMS [121] et la proposition 6 de KANNAN, NAOR et RUDICH [106], ainsi que nos résultats 17 et 18, nous obtenons directement le corollaire suivant :

Corollaire 6. Soit $\Gamma := G_1 \times \dots \times G_m$ et soit G un sous-graphe de Γ avec n sommets et de VC-dimension induite (resp. mineure) d (resp. d^*). Alors G admet un schéma d'étiquetage d'adjacence avec des étiquettes de taille :

- (1) $O(d \cdot \log n)$, si Γ est un hypercube ;
- (2) $O(\beta_0 \cdot d \cdot \log n)$, où $\beta_0 = \lceil \max\{\deg^*(\pi_1(G)), \dots, \deg^*(\pi_m(G))\} \rceil$;
- (3) $O(d^* \cdot \mu(H) \cdot \log n)$, si $G_1, \dots, G_m \in \mathcal{G}(H)$;
- (4) $O(\delta \cdot \log^2 n)$, si $G_1, \dots, G_m \in \mathcal{G}_\delta$;
- (5) $O(\log^{k+2} n)$, si $G_1, \dots, G_m \in \mathcal{G}_{(\log, k)}$;
- (6) $O(\text{dd}^*(G) \cdot d \cdot \log n)$, si G_1, \dots, G_m sont démontables ;
- (7) $O(\omega(G) \cdot d \cdot \log n)$, si G_1, \dots, G_m sont des graphes cordaux ;
- (8) $O(\omega(G) \cdot d \cdot \log n)$, si G_1, \dots, G_m sont des octaèdres.

Les assertions (1), (2), (3), (6) et (7) sont déduites du théorème 18, (4) et (5) sont dues au théorème 17 et le point (8) suit de la proposition 9.

Chapitre 6

Sous-graphes de demi-cubes

Sommaire

6.1	Résultat principal	94
6.2	Préliminaires	95
6.3	Familles d'ensembles plantées et cliques plantées	96
6.4	La clique-VC-dimension	98
6.4.1	La clique-VC-dimension des familles paires plantées	99
6.4.2	La clique-VC-dimension des familles paires quelconques	101
6.5	Preuve du théorème 19	102
6.5.1	d -décalages de familles paires plantées	102
6.5.2	Bouquets de demi-cubes	105
6.5.3	Dégénérescence des bouquets de demi-cubes	106
6.5.4	Preuve du théorème 19	107
6.6	Discussion	108

Nous nous inspirons ici de la technique de décalage de HAUSSLER [99] pour montrer que, si \mathcal{S} est une famille quelconque de sous-ensembles de $\{1, \dots, m\}$ et que $G_{1,2}(\mathcal{S}) = (V, E)$ est le graphe de 1,2-inclusion de \mathcal{S} (c.-à-d., le sous-graphe de Q_m^2 induit par \mathcal{S}), alors $\frac{|E|}{|V|} \leq \binom{d}{2}$, où $d := cVC\text{-dim}(\mathcal{S})$ dénote la clique-VC-dimension de \mathcal{S} (introduite dans le chapitre courant).

6.1 Résultat principal

Dans ce chapitre, nous présentons un analogue du théorème 8 pour les graphes de 1,2-inclusion. Nous verrons que les graphes de 1,2-inclusion peuvent contenir des cliques arbitrairement grandes pour une VC-dimension constante, il est donc nécessaire d'adapter la notion classique de VC-dimension pour capturer ce phénomène. Pour cela, nous introduisons la notion de clique-VC-dimension

$\text{cVC-dim}^*(\mathcal{S})$ d'une famille d'ensembles \mathcal{S} et nous montrons le résultat suivant :

Théorème 19. *Soit $\mathcal{S} \subseteq 2^X$ une famille de sous-ensembles d'un univers X de taille m . Soient $d := \text{cVC-dim}^*(\mathcal{S})$ la clique-VC-dimension de \mathcal{S} et $G_{1,2}(\mathcal{S}) = (V, E)$ son graphe de 1,2-inclusion. Alors, nous avons*

$$\frac{|E|}{|V|} \leq \binom{d}{2}.$$

Dans les sections 6.2 et 6.3, nous posons l'essentiel des définitions spécifiques à ce chapitre. Dans la section 6.4, nous introduisons une notion de VC-dimension adaptée à l'étude de la densité des sous-graphes de demi-cubes : la clique-VC-dimension. Dans la section 6.5, nous démontrons le théorème 19, résultat principal de ce chapitre, en adaptant la technique de preuve par décalage de HAUSSLER [99] présentée dans le chapitre 2. Enfin, dans la section 6.6, nous discutons de notre choix de VC-dimension en le mettant en relation avec d'autres manières de la définir (voir aussi la table 6.1), cette discussion donnant lieu à des questions laissées ouvertes.

6.2 Préliminaires

Soit \mathcal{S} une famille de sous-ensembles d'un univers $X = \{e_1, \dots, e_m\}$. Le graphe de 1,2-inclusion $G_{1,2}(\mathcal{S})$ de \mathcal{S} est le graphe dont l'ensemble des sommets est \mathcal{S} et dans lequel deux sommets S et S' sont reliés par une arête si et seulement si $1 \leq |S \Delta S'| \leq 2$. En d'autres mots, $G_{1,2}(\mathcal{S})$ est le sous-graphe de Q_m^2 induit par \mathcal{S} . Le graphe $G_{1,2}(\mathcal{S})$ possède à la fois les arêtes du graphe de 1-inclusion $G_1(\mathcal{S})$ de \mathcal{S} et les arêtes des deux sous-graphes de demi-cubes induits par les sommets de cardinal pairs et impairs de \mathcal{S} . Ces dernières arêtes SS' peuvent être de deux types : *verticales* lorsque $|S| = |S'| + 2$ ou $|S'| = |S| + 2$; ou *horizontales* lorsque $|S| = |S'|$.

Lorsque tous les ensembles de \mathcal{S} sont de cardinal pair, nous appelons \mathcal{S} une famille d'ensembles pairs (pour faire court, nous parlerons plutôt de *famille paire*). Dans ce cas, le graphe de 1,2-inclusion $G_{1,2}(\mathcal{S})$ coïncide avec le sous-graphe de demi-cubes $\frac{1}{2}Q_m$ induit par \mathcal{S} . En fait, Q_m^2 est isomorphe à $\frac{1}{2}Q_{m+1}$ et, par conséquent, tout graphe de 1,2-inclusion est un sous-graphe induit de demi-cube. En effet, toute famille d'ensembles $\mathcal{S} \subseteq 2^X$ peut être *augmentée* en une famille d'ensembles pairs $\mathcal{S}^+ \subset 2^{X \cup \{e_{m+1}\}}$ de telle sorte que le graphe de 1,2-inclusion de \mathcal{S} soit isomorphe à celui de \mathcal{S}^+ . Plus précisément, \mathcal{S}^+ contient tous les sous-ensembles de cardinal pair de \mathcal{S} tels quels, et tous les sous-ensembles de cardinal impair dans chacun desquels l'élément $\{e_{m+1}\}$ est

ajouté. Le lemme suivant peut aisément être démontré :

Lemme 17. *Le graphe de 1,2-inclusion $G_{1,2}(\mathcal{S})$ de toute famille d'ensembles \mathcal{S} est toujours isomorphe au graphe de 1,2-inclusion $G_{1,2}(\mathcal{S}^+)$ de sa famille augmentée \mathcal{S}^+ .*

6.3 Familles d'ensembles plantées et cliques plantées

Lorsqu'une famille d'ensembles \mathcal{S} contient l'ensemble vide \emptyset , nous qualifions cette dernière de *plantée*. Toute famille d'ensembles peut être transformée en une famille plantée par l'opération de plantation définie ci-après. Pour un ensemble $A \in \mathcal{S}$, nous dénoterons l'ensemble $\{S \Delta A : S \in \mathcal{S}\}$ par $\mathcal{S} \Delta A$. Nous dirons alors que cet ensemble $\mathcal{S} \Delta A$ est obtenu par une opération de *plantation de \mathcal{S} relativement à A* . Il apparaît clairement que cette opération de plantation est une bijection de \mathcal{S} vers $\mathcal{S} \Delta A$ qui envoie l'ensemble $A \in \mathcal{S}$ sur l'ensemble vide. Par conséquent, $\mathcal{S} \Delta A$ est une famille plantée pour tout choix de $A \in \mathcal{S}$. Nous pouvons aussi remarquer que la plantation d'une famille paire donne nécessairement une famille paire. Une fois de plus, dénotons par $G_1(\mathcal{S})$ le graphe de 1-inclusion de \mathcal{S} . De nouveau, il n'est pas difficile de démontrer que l'opération de plantation possède les propriétés suivantes :

Proposition 10. *Pour toute famille d'ensembles $\mathcal{S} \subseteq 2^X$ et pour tout ensemble $A \subseteq X$, $G_1(\mathcal{S} \Delta A) \simeq G_1(\mathcal{S})$ et $\text{VC-dim}(\mathcal{S} \Delta A) = \text{VC-dim}(\mathcal{S})$.*

Par une preuve similaire à celle de la proposition 10 (consistant essentiellement à vérifier que $|S \Delta S'| \in \{1, 2\} \Leftrightarrow |(S \Delta A) \Delta (S' \Delta A)| \in \{1, 2\}, \forall S, S' \in \mathcal{S}$), nous pouvons aussi démontrer que :

Proposition 11. *Pour toute famille d'ensembles $\mathcal{S} \subseteq 2^X$ et pour tout $A \subseteq X$, $G_{1,2}(\mathcal{S} \Delta A) \simeq G_{1,2}(\mathcal{S})$.*

Lorsqu'une clique \mathcal{C} de $\frac{1}{2}Q_m$ sera induite par une famille d'ensembles plantée, nous l'appellerons une *clique plantée*.

Lemme 18. *Toute clique \mathcal{C} dans $\frac{1}{2}Q_m$ peut être transformée en clique plantée par une opération de plantation.*

Démonstration. Soit \mathcal{C} une clique de $\frac{1}{2}Q_m$. Soit A un ensemble de cardinal maximal appartenant à \mathcal{C} . L'opération de plantation de \mathcal{C} relativement à A transforme alors \mathcal{C} en une clique plantée $\mathcal{C} \Delta A$ de $\frac{1}{2}Q_m$. En effet, si C' et C'' désignent deux sommets de \mathcal{C} , alors $|(C' \Delta A) \Delta (C'' \Delta A)| = |C' \Delta C''| = 2$. Puisque $A \Delta A = \emptyset$, $\mathcal{C} \Delta A$ est une clique plantée (voir la figure 6.1 pour une illustration). \square

Nous décrivons maintenant la structure des cliques plantées dans les demi-cubes.

Lemme 19. Une clique plantée maximale du demi-cube $\frac{1}{2}Q_m$ doit avoir l'une de ces formes :

- (a) une clique "sporadique" de taille 4 de la forme $\{\emptyset, \{e_i, e_j\}, \{e_i, e_k\}, \{e_j, e_k\}\}$ pour trois éléments quelconques $e_i, e_j, e_k \in X$;
- (b) une clique de taille m de la forme $\{\emptyset\} \cup \{\{e_i, e_j\} : e_j \in X \setminus \{e_i\}\}$, pour un élément arbitraire (mais fixé) $e_i \in X$.

Démonstration. Puisque \mathcal{C} est une clique plantée, \emptyset est l'un de ses sommets, nous le dénotons par C_0 . Tous les voisins de C_0 dans $\frac{1}{2}Q_m$ doivent être de la forme $\{e_i, e_j\}$, avec $e_i, e_j \in X$. Autrement dit, le voisinage de C_0 dans le demi-cube $\frac{1}{2}Q_m$ est le graphe adjoint ("line graph" en anglais) du graphe complet K_m dont X serait l'ensemble des sommets. En particulier, $\mathcal{C}_0 := \mathcal{C} \setminus \{C_0\}$ correspond à l'ensemble des arêtes deux-à-deux incidentes de K_m . Il n'est pas difficile de montrer que cet ensemble d'arêtes définit soit un triangle, soit une étoile dans K_m . En effet, choisissons une arête $e_i e_j$ de K_m correspondant à une paire $\{e_i, e_j\}$ de \mathcal{C}_0 . Si l'ensemble d'arêtes respectif n'est pas une étoile de K_m , alors \mathcal{C}_0 doit forcément contenir deux paires de la forme $\{e_i, e_k\}$ et $\{e_j, e_l\}$, toutes deux différentes de $\{e_i, e_j\}$. Mais alors $k = l$, car autrement les arêtes $e_i e_k$ et $e_j e_l$ ne pourraient pas être incidentes. Par conséquent, \mathcal{C}_0 contient les trois paires $\{e_i, e_j\}$, $\{e_i, e_k\}$ et $\{e_j, e_k\}$. Si \mathcal{C}_0 contient une paire supplémentaire, cette paire doit nécessairement être disjointe d'une des autres paires, ce qui n'est pas possible par hypothèse. Donc, dans ce cas, $\mathcal{C} = \{\emptyset, \{e_i, e_j\}, \{e_i, e_k\}, \{e_j, e_k\}\}$. Dans le cas contraire, si l'ensemble d'arêtes respectif est une étoile de centre e_i , alors \mathcal{C}_0 est une clique de taille $m - 1$ de la forme $\{\{e_i, e_j\} : e_j \in X \setminus \{e_i\}\}$. \square

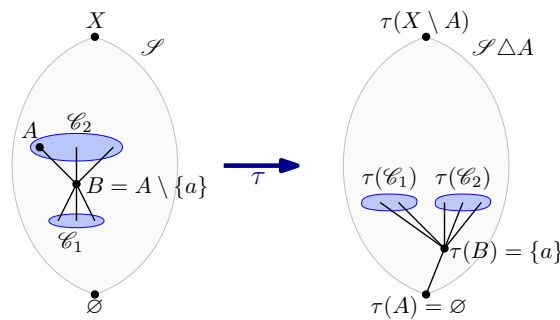


FIGURE 6.1 – Exemple d'une opération de plantation $\tau : S \mapsto S \triangle A$ d'une clique vers une clique plantée.

6.4 La clique-VC-dimension

Comme nous l'avons déjà remarqué, la VC-dimension classique d'une famille d'ensembles ne peut pas être (directement) utilisée pour borner la densité de son graphe de 1,2-inclusion. En effet, le graphe de 1,2-inclusion de la famille $\mathcal{S}_0 := \{\{e_j\} : e_j \in X\}$ est un graphe complet mais la VC-dimension de \mathcal{S}_0 vaut pourtant 1.

Nous allons donc définir une notion plus appropriée à notre contexte que nous appellerons clique-VC-dimension. L'idée ici sera de pulvériser les motifs correspondant aux cliques plantées de $\frac{1}{2}Q_m$, établis dans la section précédente (lemme 19). Rappelons que BOUSQUET et THOMASSÉ [36] ont défini une notion de 2-pulvérisation et de 2VC-dimension afin d'établir la propriété d'Erdős-Pósa pour les familles de boules d'un rayon donné r dans les graphes de 2VC-dimension bornée, et ces notions sont similaires à celles que nous allons introduire dans la mesure où elles concernent la pulvérisation non pas de tous les motifs (comme le fait la pulvérisation classique) mais seulement de certains d'entre eux (en l'occurrence, les sous-ensembles à 2 éléments de X). Rappelons qu'une famille d'ensembles \mathcal{S} 2-pulvérise un ensemble $Y \subseteq X$ si, pour tout 2-ensemble $\{e_i, e_j\} \subseteq Y$, il existe un $S \in \mathcal{S}$ tel que $Y \cap S = \{e_i, e_j\}$. La 2VC-dimension de \mathcal{S} correspond alors au cardinal maximum d'un ensemble 2-pulvérisé. Nous pouvons alors remarquer que la 2VC-dimension de \mathcal{S}_0 est égale à 0.

Soient $X = \{e_1, \dots, e_m\}$, $\mathcal{S} \subseteq 2^X$ et Y un sous-ensemble de X . Dénotons par $Q[Y]$ le sous-cube de Q_m induit par tous les sous-ensembles de Y . De même, pour un sous-ensemble $Y' \subseteq Y$, dénotons par $Q[Y', Y]$ le plus petit sous-cube de Q_m contenant Y' et Y :

$$Q[Y', Y] = \{Z \subset X : Y' \subseteq Z \subseteq Y\}.$$

En particulier, $Q[Y] = Q[\emptyset, Y]$. Pour un sommet Z de $Q[Y', Y]$, appelons

$$F(Z) := \{Z \cup Z' : Z' \subseteq X \setminus Y\}$$

la fibre de Z respectivement au cube $Q[Y', Y]$. Posons

$$\pi_{Q[Y', Y]}(\mathcal{S}) := \{Z \in Q[Y', Y] : F(Z) \cap \mathcal{S} \neq \emptyset\}$$

la projection de \mathcal{S} sur $Q[Y', Y]$. Le cube $Q[Y', Y]$ (avec $Y' \subseteq Y$) est pulvérisé par \mathcal{S} si $\pi_{Q[Y', Y]}(\mathcal{S}) = Q[Y', Y]$ (c.-à-d., pour tout $Y' \subseteq Z \subseteq Y$ la fibre $F(Z)$ contient au moins un ensemble de \mathcal{S} (voir la figure 6.2). En particulier, un sous-ensemble Y est pulvérisé par \mathcal{S} si et seulement si $\pi_{Q[Y]}(\mathcal{S}) = Q[Y]$.

6.4.1 La clique-VC-dimension des familles paires plantées

Soit \mathcal{S} une famille paire plantée de sous-ensembles d'un univers X . Soit Y un sous-ensemble de X et soit e_i un élément de X n'appartenant pas à Y . Dénotons par $P(e_i, Y)$ l'ensemble de tous les 2-ensembles contenant e_i et un élément de Y (c.-à-d., toutes les paires de 2^X de la forme $\{e_i, e_j\}$ avec $e_j \in Y$). Alors $Q[\{e_i\}, Y \cup \{e_i\}]$ constitue le plus petit sous-cube de Q_m contenant e_i et tous les 2-ensembles de $P(e_i, Y)$. Pour simplifier l'écriture, nous dénoterons ce cube par $Q(e_i, Y)$.

Nous dirons qu'une paire (e_i, Y) avec $Y \subset X$ et $e_i \in X \setminus Y$ est *c-pulvérisée* par \mathcal{S} s'il existe une fonction surjective $f : \pi_{Q(e_i, Y)}(\mathcal{S}) \rightarrow P(e_i, Y)$ telle que pour tout $S \in \pi_{Q(e_i, Y)}(\mathcal{S})$, l'inclusion $f(S) \subseteq S$ soit vérifiée. En d'autres termes, (e_i, Y) est c-pulvérisé par \mathcal{S} si chaque 2-ensemble $\{e_i, e_j\} \in P(e_i, Y)$ admet une extension $S_j \in \pi_{Q(e_i, Y)}(\mathcal{S})$ telle que $\{e_i, e_j\} \subseteq S_j$ et, pour toute paire de 2-ensembles $\{e_i, e_j\}, \{e_i, e_{j'}\} \in P(e_i, Y)$, les ensembles S_j et $S_{j'}$ soient distincts. Puisque l'ensemble vide appartient à \mathcal{S} , il est toujours c-pulvérisé par la famille.

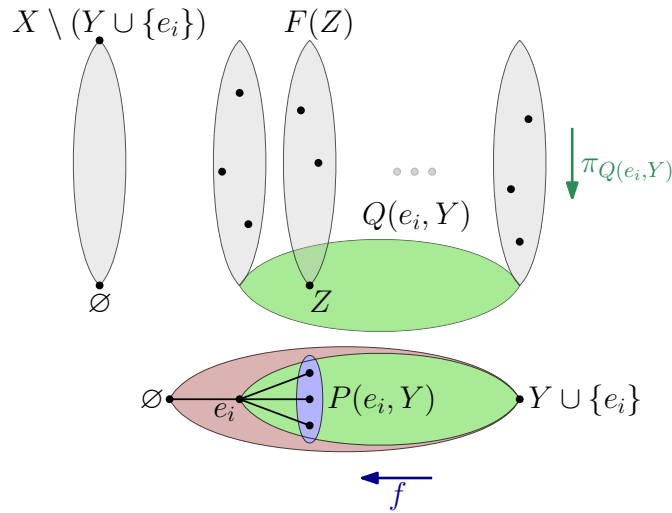


FIGURE 6.2 – Exemple d'une paire c-pulvérisée (e_i, Y) . $F(Z)$ est la fibre de Z dans $Q(e_i, Y)$. Les ensembles de \mathcal{S} (points noirs) dans les fibres des ensembles de $Q(e_i, Y)$ sont projetés sur $Q(e_i, Y)$ (en vert). Les sommets de $Q(e_i, Y)$ sont ensuite envoyés sur $P(e_i, Y)$ (en bleu) par l'opération de c-pulvérisation f . Les sommets restants de $Q[\emptyset, Y \cup \{e_i\}]$ (en rouge) ne sont "pas utilisés" pour la c-pulvérisation.

La *clique-VC-dimension (simple)* $cVC\text{-dim}(\mathcal{S})$ d'une famille paire plantée \mathcal{S} est définie de la façon suivante :

$$cVC\text{-dim}(\mathcal{S}) := 1 + \max\{|Y| : Y \subset X, \exists e_i \in X \setminus Y \text{ t.q. } \mathcal{S} \text{ c-pulvérisé } (e_i, Y)\}.$$

Remarque 19. Nous cherchons, dans la définition de la clique-VC-dimension à

c-pulvériser des “grosses cliques”. Les grosses cliques de $G_{1,2}(\mathcal{S})$ sont celles de type (b) du lemme 19 (celles de type (a) sont de taille constante égale à 4). C’est donc ce qui est fait dans la définition. Le “+1” correspond à l’ensemble vide et le “max” aux 2-ensembles de la clique.

Nous continuons en illustrant avec quelques exemples la notion clique-VC-dimension simple.

Exemple 3. Pour la famille d’ensembles $\mathcal{S}_0 = \{\{e_j\} : e_j \in X\}$ introduite plus haut, notons $\mathcal{S}_0^+ = \{\{e_j, e_{m+1}\} : e_j \in X\}$ l’augmentation de \mathcal{S}_0 vers une famille d’ensembles paire. Pour un élément arbitraire (mais fixé) e_i , posons $\mathcal{S}_1 := \{\emptyset\} \cup \{\{e_i, e_j\} : e_j \neq e_i\}$. Alors \mathcal{S}_1 coïncide avec $\mathcal{S}_0 \Delta \{e_i\}$ et avec $\mathcal{S}_0^+ \Delta \{e_i, e_{m+1}\}$. \mathcal{S}_1 est une famille paire plantée. Son graphe de 1,2-inclusion est une clique plantée et $\text{cVC-dim}(\mathcal{S}_1) = |X| = m$.

Exemple 4. Soit $\mathcal{S}_2 := \{\emptyset, \{e_1, e_2\}, \{e_1, e_3\}, \{e_2, e_3\}\}$ la clique sporadique de taille 4 introduite dans le lemme 19. Dans ce cas, il est facile de constater que l’on peut c-pulvériser n’importe quel choix de deux paires dans $\{e_1, e_2\}, \{e_1, e_3\}, \{e_2, e_3\}$, mais qu’il est impossible de c-pulvériser les trois simultanément. Cela montre que $\text{cVC-dim}(\mathcal{S}_2) = 2 + 1 = 3$.

Exemple 5. Pour deux entiers pairs quelconques m et k , dénotons par X l’univers de taille $m + km$ correspondant à l’union disjointe des $m + 1$ ensembles X_0, X_1, \dots, X_m , où $X_0 := \{e_1, \dots, e_m\}$ et $X_i := \{e_{i1}, \dots, e_{ik}\}$ pour chaque $i \in \{1, \dots, m\}$. Soit \mathcal{S}_3 la famille d’ensembles paire plantée constituée par l’ensemble vide \emptyset , l’univers X et, pour chaque $i \in \{1, \dots, m\}$, de tous les 2-ensembles de $P(e_i, X_i) = \{\{e_i, e_{i1}\}, \dots, \{e_i, e_{ik}\}\}$. Alors $G_{1,2}(\mathcal{S}_3)$ consiste en un sommet isolé X et m cliques maximales $\mathcal{C}_i := P(e_i, X_i) \cup \{\emptyset\}$ de tailles $k + 1$. De plus, ces cliques s’intersectent deux-à-deux en un unique sommet \emptyset . Nous affirmons que $\text{cVC-dim}(\mathcal{S}_3) = k + 2$. En effet, soit Y un ensemble consistant en X_i pour un certain choix de $i \in \{1, \dots, m\}$ ainsi qu’en le singleton $\{e_{(i+1)1}\}$. Alors la paire (e_i, Y) est c-pulvérisée par \mathcal{S}_3 . L’opération de c-pulvérisation $f : \pi_{Q(e_i, Y)}(\mathcal{S}_3) \rightarrow P(e_i, Y)$ est définie comme suit. Chaque 2-ensemble de $P(e_i, X_i) \subset Q(e_i, Y)$ est dans \mathcal{S}_3 et est donc envoyé sur lui-même, $X \cap (Y \cup \{e_i\}) = Y \cup \{e_i\}$ est une extension du 2-ensemble restant $\{e_i, e_{(i+1)1}\}$ dans $Q(e_i, Y)$, et donc $f(Y \cup \{e_i\}) := \{e_i, e_{(i+1)1}\}$. Puisque $|Y| = k + 1$, nous avons montré que $\text{cVC-dim}(\mathcal{S}_3) \geq k + 2$. D’autre part, $\text{cVC-dim}(\mathcal{S}_3) \leq k + 2$ car chaque élément e de X appartient à, au plus, $k + 1$ ensembles de \mathcal{S}_3 . Par conséquent, $\text{cVC-dim}(\mathcal{S}_3) = k + 2$.

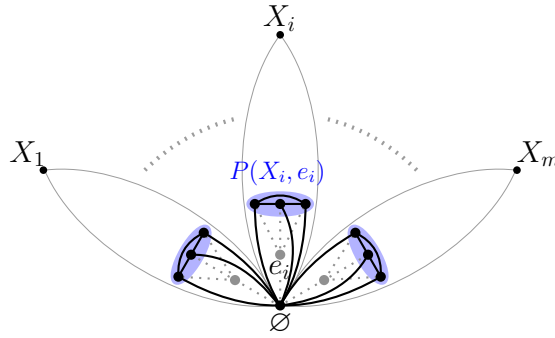


FIGURE 6.3 – Illustration de l'exemple 5.

Remarque 20. Nous pouvons remarquer que, contrairement à la clique-VC-dimension, la VC-dimension classique est insensible à l'opération de plantation. Voir, par exemple, les lignes \mathcal{S}_3 et $\mathcal{S}_3 \Delta X$ de la table 6.1.

6.4.2 La clique-VC-dimension des familles paires quelconques

La *clique-VC-dimension* $cVC\text{-dim}^*(\mathcal{S})$ d'une famille paire \mathcal{S} est le minimum sur les clique-VC-dimensions simples de toutes les familles paires plantées $\mathcal{S} \Delta A$ avec $A \in \mathcal{S}$:

$$cVC\text{-dim}^*(\mathcal{S}) := \min\{cVC\text{-dim}(\mathcal{S} \Delta A) : A \in \mathcal{S}\}.$$

La *clique-VC-dimension* $cVC\text{-dim}^*(\mathcal{S})$ d'une famille d'ensembles arbitraire \mathcal{S} est la clique-VC-dimension de son augmentation \mathcal{S}^+ .

Remarque 21. Une simple analyse des familles paires des exemples 3-5 montre que $cVC\text{-dim}^*(\mathcal{S}_1) = m$, $cVC\text{-dim}^*(\mathcal{S}_2) = 3$ et $cVC\text{-dim}^*(\mathcal{S}_3) = k + 2$.

Remarque 22. La famille d'ensembles \mathcal{S}_3 montre que le degré maximum du graphe de 1,2-inclusion $G_{1,2}(\mathcal{S})$ d'une famille paire \mathcal{S} peut être arbitrairement plus grand que sa clique-VC-dimension $cVC\text{-dim}^*(\mathcal{S})$. En effet, \emptyset est le sommet de plus gros degré de $G_{1,2}(\mathcal{S}_3)$ et son degré est égal à km .

La famille \mathcal{S}_3 justifie aussi que nous définissions la clique-VC-dimension de \mathcal{S} comme le minimum sur tous les $\mathcal{S} \Delta A$ avec $A \in \mathcal{S}$. En effet, il est important d'insister ici sur le lemme 11 : l'opération de plantation définit un isomorphisme de graphe. En d'autres termes, la famille d'ensembles \mathcal{S} est (lourdement) affectée par cette application (de même que sa clique-VC-dimension simple), mais la densité de son graphe de 1,2-inclusion G ne l'est pas. Nous pourrions donc supposer que G est représenté par une famille de sous-ensembles de X atteignant la plus petite clique-VC-dimension simple. C'est essentiellement ce qui est fait en choisissant un "min" sur toutes les plantations de \mathcal{S} dans la définition de $cVC\text{-dim}^*$. Considérons

la plantation de \mathcal{S}_3 respectivement à l'ensemble $X \in \mathcal{S}_3$. Alors nous pouvons remarquer que $\text{cVC-dim}(\mathcal{S}_3 \Delta X) \geq (m-1)k + 1$ car

$$\mathcal{S}_3 \Delta X = \{\emptyset, X\} \cup \left(\bigcup_{(i,j) \in \{1, \dots, m\} \times \{1, \dots, k\}} \{X \setminus \{e_i, e_{ij}\}\} \right).$$

Soit $Y := \{e_{ij} : i \in \{1, \dots, m-1\} \text{ et } j \in \{1, \dots, k\}\}$. Nous affirmons que (e_1, Y) est c-pulvérisé par $\mathcal{S} \Delta X$. Pour montrer cela, posons $\mathcal{S}'_3 := \pi_{Q(e_1, Y)}(\mathcal{S}_3 \Delta X)$, $S_{ij} := X \setminus \{e_i, e_{ij}\}$ et $S'_{ij} := \pi_{Q(e_1, Y)}(S_{ij})$. Soit $f : \mathcal{S}'_3 \rightarrow P(e_1, Y)$ une application telle que, pour tout $i \in \{2, \dots, m\}$ et $j \in \{1, \dots, k\}$, $f(S'_{ij}) = \{e_1, e_{(i-1)j}\}$. Alors chaque $\{e_1, e_{(i-1)j}\}$ possède clairement une extension S'_{ij} avec une fibre non vide ($S_{ij} \in F(S'_{ij})$). Nous avons qui plus est que, pour tout $S_{rl} \neq S_{ij}$, $S'_{rl} \neq S'_{ij}$. Par conséquent, f est surjective et nous pouvons conclure que (e_1, Y) est c-pulvérisée. Puisque $|Y| = (m-1)k$, nous avons $\text{cVC-dim}(\mathcal{S}_3 \Delta X) \geq (m-1)k + 1$.

6.5 Preuve du théorème 19

Maintenant posées les remarques et définitions des sections précédentes, nous pouvons avancer sur la preuve du résultat principal. Nous commençons la preuve par la définition du double décalage (d-décalage) qui sera notre adaptation du décalage classique, introduit dans la preuve de HAUSSLER [99] (voir chapitre 2), aux familles paires plantées. Nous montrons que l'opération de d-décalages satisfait les mêmes propriétés essentielles (1), (2) et (3) que le décalage classique et que le résultat d'une séquence complète de d-décalages est toujours un bouquet de demi-cubes (famille plantée paire particulière que nous définirons plus loin). Nous montrons que la dégénérescence du graphe de 1,2-inclusion de tels bouquets \mathcal{B} est bornée par $\binom{d}{2}$, où $d := \text{cVC-dim}(\mathcal{B})$. Nous concluons la preuve du théorème en considérant des familles paires \mathcal{S} quelconques et en appliquant les arguments précédents aux familles paires plantées $\mathcal{S} \Delta A$, avec A un ensemble de \mathcal{S} tel que $\text{cVC-dim}(\mathcal{S} \Delta A) = \text{cVC-dim}^*(\mathcal{S})$.

6.5.1 d-décalages de familles paires plantées

Un *double décalage* (ou *d-décalage*) d'une famille paire plantée $\mathcal{S} \subseteq 2^X$ respectivement à un 2-ensemble $\{e_i, e_j\} \in 2^X$ est une application $\varphi_{ij} : \mathcal{S} \rightarrow 2^X$ remplaçant chaque ensemble S de \mathcal{S} incluant $\{e_i, e_j\}$ et tel que $S \setminus \{e_i, e_j\} \notin \mathcal{S}$

par l'ensemble $S \setminus \{e_i, e_j\}$:

$$\begin{aligned} \varphi_{ij} : \mathcal{S} &\rightarrow 2^X \\ S &\mapsto \begin{cases} S \setminus \{e_i, e_j\}, & \text{if } \{e_i, e_j\} \subseteq S \text{ et } S \setminus \{e_i, e_j\} \notin \mathcal{S} \\ S, & \text{sinon.} \end{cases} \end{aligned}$$

Lemme 20. Soit $\mathcal{S} \subseteq 2^X$ une famille paire plantée, soit $\{e_i, e_j\} \subseteq X$ un 2-ensemble, et soit $G_{1,2}(\mathcal{S}) = G = (V, E)$ et $G_{1,2}(\varphi_{ij}(\mathcal{S})) = G' = (V', E')$ les graphes de 1,2-inclusion respectifs de \mathcal{S} et de $\varphi_{ij}(\mathcal{S})$. Alors $|V| = |V'|$ et $|E| \leq |E'|$.

Démonstration. Le fait qu'un d-décalage φ_{ij} préserve le nombre de sommets d'un sous-graphe induit de demi-cube est évident d'après sa définition. Il nous suffit donc de démontrer que φ_{ij} ne peut pas diminuer le nombre d'arêtes. Pour cela, nous allons étendre φ_{ij} à une application injective $\psi_{ij} : E \rightarrow E'$. Nous dirons qu'une arête SS' de G est *stable* si $\varphi_{ij}(S) = S$ et $\varphi_{ij}(S') = S'$. Dans le cas contraire, nous dirons que l'arête est *décalable*. Chaque arête stable SS' est sa propre image par ψ_{ij} (c.-à-d., $\psi_{ij}(SS') := SS'$).

Considérons à présent une arête décalable SS' de E . Remarquons que dans ce cas, $\{e_i, e_j\} \subseteq S$ ou $\{e_i, e_j\} \subseteq S'$. Afin de définir $\psi_{ij}(SS')$, nous devons distinguer deux cas selon que $\{e_i, e_j\}$ ne soit un sous-ensemble que d'une extrémité (S ou bien S') de l'arête ou bien des deux.

Cas 1'. $\{e_i, e_j\} \subseteq S$ et $\{e_i, e_j\} \not\subseteq S'$ (le cas $\{e_i, e_j\} \subseteq S'$ et $\{e_i, e_j\} \not\subseteq S$ est équivalent à un renommage près de S et S').

Étant donné que $\{e_i, e_j\} \not\subseteq S'$, nous avons forcément $\varphi_{ij}(S') = S'$. Puisque SS' est décalable, $\varphi_{ij}(S) \neq S$ (c.-à-d., $\varphi_{ij}(S) = S \setminus \{e_i, e_j\} =: Z$). Nous devons à nouveau considérer deux cas, selon qu'un unique élément (e_i ou bien e_j) appartienne à S' ou bien les deux.

Sous-cas 1'.1. $e_i \in S'$ et $e_j \notin S'$ (le cas $e_j \in S'$ et $e_i \notin S'$ est similaire). Dans ce cas, il existe un élément $e_k \in X$ tel que $S \Delta S' = \{e_j, e_k\}$. Remarquons que $S \not\subseteq S'$ car $e_j \notin S'$ et $e_j \in S$. Par conséquent, ou bien $S' \subseteq S$, ou bien il existe un $A \subset X$ tel que $S' = A \cup \{e_k\}$ et $S = A \cup \{e_j\}$. Dans le premier cas, nous avons $S = S' \cup \{e_j, e_k\}$, $Z = S' \cup \{e_k\} \setminus \{e_i\}$ et $Z \Delta S' = \{e_i, e_k\}$. Dans le second cas, nous avons $Z = A \setminus \{e_i\}$ et $Z \Delta S' = \{e_i, e_k\}$. Quoi qu'il en soit, $|Z \Delta S'| = 2$ et $ZS' \in E'$. Nous posons alors $\psi_{ij}(SS') := ZS'$.

Sous-cas 1'.2. $e_i \notin S'$ et $e_j \notin S'$. Alors $S \Delta S' = \{e_i, e_j\}$ et donc $S \setminus \{e_i, e_j\} = Z = S'$. Mais nous obtenons alors une contradiction avec le fait que SS' soit décalable (c.-à-d., $Z = \varphi_{ij}(S)$ ne peut pas appartenir à \mathcal{S}).

Cas 2'. $\{e_i, e_j\} \subseteq S$ et $\{e_i, e_j\} \subseteq S'$.

Posons $Z := S \setminus \{e_i, e_j\}$ et $Z' := S' \setminus \{e_i, e_j\}$. Alors Z et Z' appartiennent tous deux à $\varphi_{ij}(\mathcal{S})$ et ZZ' définit une arête de G' . Puisque SS' est décalable, au moins l'un des ensembles Z ou Z' n'appartient pas à \mathcal{S} .

Sous-cas 2'.1. $Z, Z' \notin \mathcal{S}$. Alors $\varphi_{ij}(S) = Z$ et $\varphi_{ij}(S') = Z'$ et ZZ' est une arête de G' . Dans ce cas, nous posons $\psi_{ij}(SS') := ZZ'$.

Sous-cas 2'.2. $Z \in \mathcal{S}$ et $Z' \notin \mathcal{S}$ (le cas $Z \notin \mathcal{S}$ et $Z' \in \mathcal{S}$ est similaire). Alors $\varphi_{ij}(S) = S$, $\varphi_{ij}(S') = Z'$ et ZZ' est une arête de G' mais pas de G . Dans ce cas, nous posons $\psi_{ij}(SS') := ZZ'$.

Il nous reste maintenant à démontrer que l'application $\psi_{ij} : E \rightarrow E'$ est injective. Supposons, par l'absurde, que G' contienne une arête ZZ' qui soit l'image de deux arêtes distinctes SS' et CC' de G (c.-à-d., $\psi_{ij}(SS') = \psi_{ij}(CC') = ZZ'$). Puisqu'au moins l'une des arêtes SS' et CC' est distincte de ZZ' , nous pouvons conclure d'après la définition d'un d-décalage que ZZ' n'est pas une arête de G . Disons $Z' \notin \mathcal{S}$. Cela implique aussi que SS' et CC' sont toutes deux décalables.

Cas 1''. $Z \notin \mathcal{S}$.

De par la définition de l'application ψ_{ij} , et puisque $Z, Z' \notin \mathcal{S}$, les deux arêtes SS' et CC' se trouvent dans le sous-cas 2'.1. Cela montre que $Z = S \setminus \{e_i, e_j\}$, $Z' = S' \setminus \{e_i, e_j\}$ et $Z = C \setminus \{e_i, e_j\}$, $Z' = C' \setminus \{e_i, e_j\}$. Cela conduit au fait que $S = C$ et $S' = C'$. C'est une contradiction avec l'hypothèse $SS' \neq CC'$.

Cas 2''. $Z \in \mathcal{S}$.

À un renommage approprié près des ensembles S , S' , C et C' , nous pouvons supposer que $\varphi_{ij}(S) = \varphi_{ij}(C) = Z$ et que $\varphi_{ij}(S') = \varphi_{ij}(C') = Z'$. Puisque $Z' \notin \mathcal{S}$, nous déduisons directement de la définition de ψ_{ij} que $S' = Z' \cup \{e_i, e_j\} = C'$. Aussi, puisque $Z \in \mathcal{S}$, nous avons soit $S = C = Z$, ce qui contredit le choix de $SS' \neq CC'$, soit $S \setminus \{e_i, e_j\} = C = Z$ (ou la possibilité symétrique $C \setminus \{e_i, e_j\} = S = Z$), ce qui contredit le fait que SS' (ou CC') soit décalable.

Cela montre que l'application $\psi_{ij} : E \rightarrow E'$ est injective, et nous permet de conclure que $|E| \leq |E'|$. \square

Lemme 21. Si φ_{ij} dénote une application de d-décalage d'une famille paire plantée $\mathcal{S} \subset 2^X$, alors $\text{cVC-dim}(\varphi_{ij}(\mathcal{S})) \leq \text{cVC-dim}(\mathcal{S})$.

Démonstration. Soit (e, Y) une paire c-pulvérisée par $\mathcal{S}_{ij} := \varphi_{ij}(\mathcal{S})$ (rappelons que $Y \subset X$ et que $e \notin Y$). Soient $\mathcal{S}' := \pi_{Q(e, Y)}(\mathcal{S})$ et $\mathcal{S}'_{ij} := \pi_{Q(e, Y)}(\varphi_{ij}(\mathcal{S}))$. De par la définition de la c-pulvérisation, il existe une application surjective f associant chaque élément de \mathcal{S}'_{ij} à un 2-ensemble $\{e, e'\} \in P(e, Y)$. Nous allons

définir une fonction g de \mathcal{S}' vers \mathcal{S}'_{ij} et l'utiliser pour construire une opération de c-pulvérisation $f' := f \circ g$ de \mathcal{S}' vers $P(e, Y)$. Soit $S_{e'} \in \mathcal{S}'_{ij}$ un ensemble tel que $f(S_{e'}) = \{e, e'\}$. Si $S_{e'}$ appartient à \mathcal{S}' , alors le 2-ensemble $\{e, e'\}$ possède aussi une extension dans \mathcal{S}' et nous pouvons poser $g(S_{e'}) := S_{e'}$. Si $S_{e'}$ n'appartient pas à \mathcal{S}' , cela signifie qu'il existe un ensemble $S \in \mathcal{S}$ tel que $S \neq \varphi_{ij}(S)$, et $\varphi_{ij}(S)$ est dans la fibre $F(S_{e'})$ de $S_{e'}$ respectivement à $\pi_{Q(e, Y)}$ dans \mathcal{S}'_{ij} . L'ensemble S est dans la fibre $F(S')$ d'un certain ensemble $S' \in \mathcal{S}'$ respectivement à $\pi_{Q(e, Y)}$. Puisque $\varphi_{ij}(S) \subseteq S$, nous avons que $S_{e'} \subseteq S'$ et que $S' \in \mathcal{S}'$ est une extension du 2-ensemble $\{e, e'\}$. Nous pouvons donc poser $g(S') := S_{e'}$. De plus, pour chaque ensemble $S' \in \mathcal{S}' \setminus \mathcal{S}'_{ij}$, il existe un ensemble $S \in F(S')$ tel que $\varphi_{ij}(S) \neq S$. Dans ce cas, il existe un ensemble $S_{e'} \in \mathcal{S}'_{ij}$ tel que $\varphi_{ij}(S) \in F(S_{e'})$. Nous définissons alors $g(S') := S_{e'}$. Comme $\varphi_{ij}(S) \subset S$, nous avons $S_{e'} \subseteq S'$.

La fonction g est surjective par définition et elle envoie chaque ensemble de \mathcal{S}' soit sur lui-même, soit sur un sous-ensemble de lui-même. Puisque f est une opération de c-pulvérisation, $f' := f \circ g$ en est aussi une et (e, Y) est c-pulvérisée par \mathcal{S} . En conclusion, $\text{cVC-dim}(\varphi_{ij}(\mathcal{S})) \leq \text{cVC-dim}(\mathcal{S})$ car chaque (e, Y) c-pulvérisée par \mathcal{S}'_{ij} est aussi c-pulvérisée par \mathcal{S} . \square

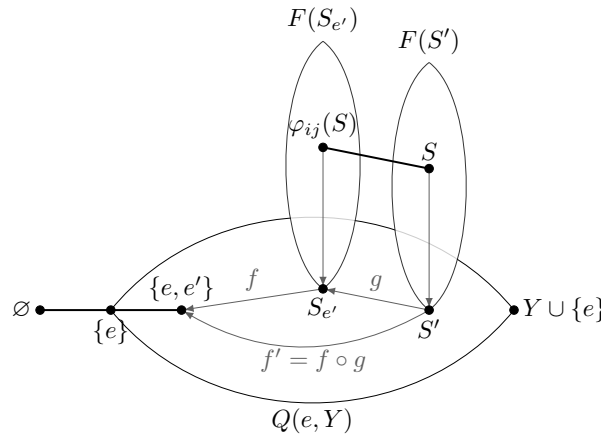


FIGURE 6.4 – Pour la preuve du lemme 21.

6.5.2 Bouquets de demi-cubes

Rappelons qu'un bouquet de demi-cubes est une famille d'ensembles paire $\mathcal{B} \subseteq 2^X$ telle que tout sous-ensemble de cardinal pair S' d'un ensemble S dans \mathcal{B} appartient aussi à \mathcal{B} .

Lemme 22. *Toute famille paire plantée $\mathcal{S} \subseteq 2^X$ peut être transformée en un bouquet de demi-cubes par une séquence finie de d -décalages.*

Démonstration. Soit $\mathcal{S}^0, \mathcal{S}^1, \mathcal{S}^2, \dots$ une séquence de familles paires plantées telle que $\mathcal{S}^0 = \mathcal{S}$ et, pour tout $i \geq 1$, $\mathcal{S}^i \neq \mathcal{S}^{i-1}$. Cette séquence est forcément finie car chaque d-décalage diminue strictement la somme des cardinaux des ensembles dans la famille. Dénotons par \mathcal{S}^r la dernière famille dans la séquence. Autrement dit, effectuer un d-décalage de \mathcal{S}^r par rapport à n'importe quelle paire d'éléments de X conduit à la même famille d'ensembles \mathcal{S}^r . Alors, pour tout $S \in \mathcal{S}^r$, et pour toute paire $\{e_j, e_k\} \subseteq S$, l'ensemble $S \setminus \{e_j, e_k\}$ appartient à \mathcal{S}^r . C'est la définition d'un bouquet de demi-cubes \mathcal{S}^r . \square

Continuons avec des propriétés simples mais essentielles des bouquet de demi-cubes.

Lemme 23. *Soit $\mathcal{B} \subset 2^X$ un bouquet de demi-cubes de clique-VC-dimension $d := \text{cVC-dim}(\mathcal{B})$. Alors les propriétés suivantes sont vérifiées :*

- (i) *pour tout élément $e_i \in X$, $|\{\{e_i, e_j\} \in \mathcal{B} : e_j \in X \setminus \{e_i\}\}| \leq d - 1$;*
- (ii) *si S est un ensemble dans \mathcal{B} , alors $|S| \leq d$;*
- (iii) *si S est un ensemble maximal par inclusion dans \mathcal{B} , alors $\mathcal{B} \setminus \{S\}$ reste un bouquet de demi-cubes.*

Démonstration. L'inégalité $|\{\{e_i, e_j\} \in \mathcal{B} : e_j \in X \setminus \{e_i\}\}| \leq d - 1$ découle directement de la définition de $\text{cVC-dim}(\mathcal{B})$. La propriété (iii) est aussi immédiate de par la définition d'un bouquet de demi-cubes. Pour démontrer la propriété (ii), faisons l'hypothèse absurde que $|S| > d$. Puisque \mathcal{B} est un bouquet de demi-cubes, chaque sous-ensemble de S de cardinal pair appartient à \mathcal{B} . Ainsi, si nous choisissons un élément $e \in S$ et si nous posons $Y := S \setminus \{e\}$, alors tous les 2-ensembles de la forme $\{e, e'\}$ avec $e' \in Y$ sont des sous-ensembles de S et, de fait, sont aussi des ensembles dans \mathcal{B} . Par conséquent, (e, Y) est c-pulvérisé par \mathcal{B} . Étant donné que $|Y| = |S| - 1 > d - 1$, cela contredit l'hypothèse $d = \text{cVC-dim}(\mathcal{B})$. \square

6.5.3 Dégénérescence des bouquets de demi-cubes

Dans cette sous-section, nous démontrons la borne supérieure suivante sur la dégénérescence des graphes de 1,2-inclusion des bouquets de demi-cubes :

Lemme 24. *Soit $\mathcal{B} \subset 2^X$ un bouquet de demi-cubes de clique-VC-dimension $d := \text{cVC-dim}(\mathcal{B})$, et soit $G := G_{1,2}(\mathcal{B})$ son graphe de 1,2-inclusion. Alors la dégénérescence de G est d'au plus $\binom{d}{2}$.*

Démonstration. Soit S un ensemble de cardinal maximal dans \mathcal{B} . Par le lemme 23(iii), $\mathcal{B} \setminus \{S\}$ est un bouquet de demi-cubes. Il suffit donc de montrer que le degré de S dans G est borné (supérieurement) par $\binom{d}{2}$. Par le lemme 23(ii), nous savons que $|S| \leq d$. Cela implique que S est incident dans G à, au plus, $\binom{d}{2}$ arêtes

verticales. Il nous reste donc à borner le nombre d'arêtes horizontales touchant S . Le résultat suivant sera utile à cette fin :

Affirmation 24.1. *Si $|S| = d - k \leq d$, alors S est incident dans G à, au plus, $(d - k)k$ arêtes horizontales.*

Preuve de l'affirmation. Choisissons un élément quelconque $s \in S$ et posons $Y := S \setminus \{s\}$. Pour un élément $e \in X \setminus S$, posons $S_s^e := Y \cup \{e\}$. Remarquons que de tels S_s^e sont exactement les voisins de S qui lui sont reliés par une arête horizontale dans $\frac{1}{2}Q_m$. Soit $X' = \{e \in X \setminus S : S_s^e \in \mathcal{B}\}$.

Choisissons un élément quelconque $y \in Y$. Alors $y \in S_s^e$ pour tout $e \in X'$. Puisque \mathcal{B} est un bouquet de demi-cubes, chacune des $d - k - 1$ paires $\{y, e'\}$ avec $e' \in S \setminus \{y\}$ appartient à \mathcal{B} . Nous en déduisons que $P(y, S \setminus \{y\}) \subseteq \mathcal{B}$. À chaque ensemble S_s^e , $e \in X'$, correspond l'unique paire $\{y, e\}$, or $\{y, e\} \in \mathcal{B}$ car $y, e \in S_s^e$. Par conséquent, $P(y, X') \subset \mathcal{B}$. Comme $|P(y, S \setminus \{y\})| + |P(y, X')| \leq d - 1$ et comme $|P(y, S \setminus \{y\})| = d - k - 1$, $|X'| = |P(y, X')|$, nous pouvons conclure que $|X'| \leq k$. Ainsi, pour un élément fixé $s \in S$, S possède au plus k voisins de la forme S_s^e avec $e \in X'$. Étant donné qu'il y a $|S| = d - k$ choix possibles pour l'élément s , S a au plus $(d - k)k$ voisins de cardinal $|S|$. ■

Nous pouvons maintenant continuer la preuve du lemme 24. Soit $|S| = d - k \leq d$. Alors S possède $\binom{d-k}{2}$ voisins de la forme $S \setminus \{e, e'\}$ avec $e \neq e' \in S$ (c.-à-d., S est incident à $\binom{d-k}{2}$ arêtes verticales). Il nous reste à borner le nombre de voisins de S de la forme $S \setminus \{e\} \cup \{e'\}$ avec $e \in S$ et $e' \in X \setminus S$. De par l'affirmation 24.1, S possède au plus $(d - k)k$ tels voisins. En résumé, S possède $(d - k)k + \binom{d-k}{2} = \frac{1}{2}(d^2 - d - k^2 + k)$ voisins dans G , et ce nombre est maximal pour $k = 0$ car

$$\frac{1}{2}(d^2 - d - k^2 + k) = \frac{1}{2}(d^2 - d) - \frac{1}{2}(k^2 - k) = \binom{d}{2} - \binom{k}{2} \leq \binom{d}{2}.$$

Finalement, le degré de S dans G est majoré par $\binom{d}{2}$, comme affirmé. □

6.5.4 Preuve du théorème 19

Soit \mathcal{S} une famille d'ensembles paire sur un univers X de cardinal m . Posons $d := \text{cVC-dim}^*(\mathcal{S})$ la clique-VC-dimension de cette famille et $G_{1,2}(\mathcal{S})$ son graphe de 1,2-inclusion. Nous devons prouver que $\frac{|E|}{|V|} \leq \binom{d}{2} =: D$.

Soit A un ensemble dans \mathcal{S} tel que $\text{cVC-dim}(\mathcal{S} \Delta A) = \text{cVC-dim}^*(\mathcal{S}) = d$. Par la proposition 11, $G_{1,2}(\mathcal{S} \Delta A) \simeq G_{1,2}(\mathcal{S})$. Il nous suffit donc de montrer que $\frac{|E(G_{1,2}(\mathcal{S} \Delta A))|}{|V(G_{1,2}(\mathcal{S} \Delta A))|} \leq D$. Considérons une séquence complète de d -décalages

de $\mathcal{S} \Delta A$ et dénotons par $(\mathcal{S} \Delta A)^*$ la famille d'ensembles résultante. Comme $\mathcal{S} \Delta A$ est une famille paire plantée, nous pouvons appliquer le lemme 21 à chaque d-décalage pour déduire que $\text{cVC-dim}((\mathcal{S} \Delta A)^*) \leq \text{cVC-dim}(\mathcal{S} \Delta A) = d$. Par le lemme 22, $(\mathcal{S} \Delta A)^*$ est un bouquet de demi-cubes et, par le lemme 24, la dégénérescence de son graphe de 1,2-inclusion $G_{1,2}((\mathcal{S} \Delta A)^*) =: G^* = (V^*, E^*)$ est d'au plus D . Ainsi, $\frac{|E^*|}{|V^*|} \leq D$ par la proposition 2. En appliquant le lemme 20 à chaque d-décalage et en prenant en compte que $G_{1,2}(\mathcal{S} \Delta A) \simeq G_{1,2}(\mathcal{S})$, nous concluons que $\frac{|E|}{|V|} \leq \frac{|E^*|}{|V^*|}$, ce qui nous conduit à l'inégalité annoncée $\frac{|E|}{|V|} \leq D$ et termine la preuve du théorème 19 dans le cas des familles d'ensembles paires. Maintenant, si \mathcal{S} est une famille d'ensembles quelconque, alors $\text{cVC-dim}^*(\mathcal{S}) = \text{cVC-dim}^*(\mathcal{S}^+)$, où \mathcal{S}^+ dénote l'augmentation de \mathcal{S} vers une famille d'ensembles paire. Comme par le lemme 17, \mathcal{S} et \mathcal{S}^+ ont des graphes de 1,2-inclusion isomorphes, l'inégalité sur la densité de \mathcal{S} découle directement de l'inégalité sur la densité de \mathcal{S}^+ . Cela conclut la preuve du théorème 19. \square

Exemple 6. De même que pour le cas classique de la VC-dimension avec le théorème 8, l'inégalité du théorème 19 entre la densité du graphe de 1,2-inclusion $G_{1,2}(\mathcal{S})$ et la clique-VC-dimension de \mathcal{S} est serrée du point de vue suivant : il existe des familles paires \mathcal{S} telles que la dégénérescence de $G_{1,2}(\mathcal{S})$ soit égale à $\binom{d}{2}$. Par exemple, la clique sporadique \mathcal{S}_2 a une clique-VC-dimension égale à 3 (voir l'exemple 4 et la remarque 21), dégénérescence 3 et densité $\frac{3}{2}$. Nous pouvons remarquer que $G_{1,2}(\mathcal{S}_2)$ correspond au demi-cube $\frac{1}{2}Q_3$. Plus généralement, considérons la famille paire \mathcal{S}_4 constituée de tous les sous-ensembles de cardinaux pairs d'un univers X de taille m . Alors nous avons clairement que $d := \text{cVC-dim}^*(\mathcal{S}_4) = |X| = m$ et \mathcal{S}_4 induit le demi-cube $\frac{1}{2}Q_m$. Nous affirmons alors que $\frac{1}{2}Q_m$ a une dégénérescence égale à $\binom{d}{2}$. En effet, chaque $S \in \mathcal{S}_4$ est incident dans $G_{1,2}(\mathcal{S}_4)$ à $\binom{|X|-|S|}{2}$ sur-ensembles de cardinal $|S| + 2$, à $\binom{|S|}{2}$ sous-ensembles de cardinal $|S| - 2$, et à $|S|(|X| - |S|)$ ensembles de cardinal $|S|$. Posons $s := |S|$. Nous pouvons maintenant conclure que S a un degré de

$$\frac{(m-s)(m-s-1) + s(s-1)}{2} + s(m-s) = \frac{1}{2}(m^2 - m) = \frac{1}{2}(d^2 - d) = \binom{d}{2}.$$

6.6 Discussion

De manière analogue à la 2-pulvérisation et à la 2VC-dimension, nous pouvons définir les concepts de star-pulvérisation et de star-VC-dimension qui pourraient éventuellement être utiles pour trouver une borne supérieure plus serrée (que celle que nous fournissons ici) sur la densité des graphes de 1,2-inclusion. Soient $Y \subset X$ et $e \notin Y$. Nous dirons qu'une famille \mathcal{S} star-pulvérise (ou, pour faire court, *s-pulvérise*) une paire (e, Y) si, pour tout $y \in Y$, il existe un ensemble

$S \in \mathcal{S}_{ij}$ tel que $S \cap (Y \cup \{e\}) = \{e, y\}$. La *star-VC-dimension simple* d'une famille d'ensembles plantée \mathcal{S} est alors définie par

$$\text{sVC-dim}(\mathcal{S}) := 1 + \max\{|Y| : Y \subset X, \exists e_i \in X \setminus Y \text{ t.q. } \mathcal{S} \text{ s-pulvérise } (e_i, Y)\}.$$

Notons que, à la différence de la c-pulvérisation, la s-pulvérisation ne nécessite pas d'introduire une application surjective $f : \pi_{Q(e,Y)}(\mathcal{S}) \rightarrow P(e, Y)$. Une paire (e, Y) est s-pulvérisée si tous les 2-ensembles de $P(e, Y)$ ont une fibre non-vide (c.-à-d., si $P(e, Y) \subseteq \pi_{Q(e,Y)}(\mathcal{S})$). Par conséquent, toute paire s-pulvérisée (e, Y) est aussi c-pulvérisée, et donc $\text{sVC-dim}(\mathcal{S}) \leq \text{cVC-dim}(\mathcal{S})$.

Puisque $\text{sVC-dim}(\mathcal{S}_3 \Delta X) = 3$ et $G_{1,2}(\mathcal{S}_3 \Delta X)$ contient une clique de taille $k + 1$, $\text{sVC-dim}(\mathcal{S})$ ne peut pas être utilisée directement pour borner la densité des graphes de 1,2-inclusion. Mais, ici encore, nous pouvons adapter cette notion en prenant la valeur maximale obtenue sur toutes les plantations par rapport aux ensembles dans \mathcal{S} . La *star-VC-dimension* $\text{sVC-dim}^*(\mathcal{S})$ d'une famille d'ensembles quelconque \mathcal{S} vaut alors $\max\{\text{sVC-dim}(\mathcal{S} \Delta A) : A \in \mathcal{S}\}$ ¹. De façon assez surprenante, bien que l'inégalité $\text{sVC-dim}(\mathcal{S}) \leq \text{cVC-dim}(\mathcal{S})$ soit vérifiée pour les familles d'ensembles plantées, il n'y a aucune relation entre $\text{cVC-dim}^*(\mathcal{S})$ et $\text{sVC-dim}^*(\mathcal{S})$. C'est ce que montrent les deux exemples suivants :

Exemple 7. Soit $X := \{1, 2, \dots, 2m - 1, 2m\}$ avec m un entier pair quelconque, et soit $\mathcal{S}_5 := \{\emptyset\} \cup \{\{1, 2, \dots, 2i - 1, 2i\} : i = 1, \dots, m\}$. Les ensembles non-vides de \mathcal{S}_5 peuvent être interprétés comme les intervalles de longueur paire de \mathbb{N} avec une origine commune. Le graphe de 1,2-inclusion de \mathcal{S}_5 est un chemin de longueur m . Pour chaque ensemble $\{1, 2, \dots, 2i\}$, la famille plantée $\mathcal{S}_5^i := \mathcal{S}_5 \Delta \{1, 2, \dots, 2i\}$ correspond à l'union des familles $\mathcal{S}' := \{\emptyset, \{2i+1, 2i+2\}, \dots, \{2i+1, 2i+2, \dots, 2m\}\}$ et $\mathcal{S}'' := \{\{1, 2, \dots, 2i-1, 2i\}, \dots, \{2i-1, 2i\}\}$. Nous allons montrer que, pour tout $i \in \{1, \dots, m\}$, nous avons $\text{sVC-dim}(\mathcal{S}_5^i) \leq 3$ et $\text{cVC-dim}(\mathcal{S}_5^i) = \max\{i, m-i\} + 1$. En effet, pour chaque élément $j \in X$, \mathcal{S}_5^i ne peut pas simultanément s-pulvériser deux paires $\{j, l_1\}, \{j, l_2\}$ avec $j < l_1 < l_2$ car chaque ensemble de \mathcal{S}_5^i contenant l_2 doit aussi contenir l_1 . D'autre part, une analyse similaire nous permet d'affirmer que \mathcal{S}_5^i ne peut pas s-pulvériser deux paires $\{j, l_1\}$ et $\{j, l_2\}$ avec $l_2 < l_1 < j$. Par conséquent, si la paire (j, Y) est s-pulvérisée par \mathcal{S}_5^i , alors $|Y| \leq 2$. Cela démontre que $\text{sVC-dim}^*(\mathcal{S}_5) \leq 3$.

Pour voir que $\text{cVC-dim}(\mathcal{S}_5^i) = \max\{i, m-i\} + 1$, remarquons tout d'abord que \mathcal{S}' c-pulvérise la paire $(2i+1, Y')$ avec $Y' := \{2i+2, 2i+4, \dots, 2m\}$ et que \mathcal{S}'' c-pulvérise la paire $(2i, Y'')$ avec $Y'' := \{1, 3, \dots, 2i-1\}$. Comme le minimum sur tous les $i \in \{1, \dots, m\}$ de $\max\{i, m-i\} + 1$ est atteint pour $i = \frac{m}{2}$, nous pouvons conclure que $\text{cVC-dim}^*(\mathcal{S}_5) = \frac{m}{2} + 1$. Au final, $\text{sVC-dim}^*(\mathcal{S})$ peut être arbitrairement inférieure à $\text{cVC-dim}^*(\mathcal{S})$.

1. Comme remarqué par un lecteur de « On density of subgraphs of halved cubes » [59] et par O. Bousquet, la star-VC-dimension moins 1 coïncide avec la notion de *star number* étudiée dans le contexte de l'apprentissage actif (HANNEKE et YANG [96, Définition 2]).

Exemple 8. Soit $X := X_1 \dot{\cup} X_2$ avec $X_1 := \{e_1, \dots, e_m\}$ et $X_2 := \{x_1, \dots, x_m\}$, et soit $\mathcal{S}_6 := \{\emptyset, \{e_1, x_1\}\} \cup \{\{e_1, e_i, x_1, x_i\} : 2 \leq i \leq m\}$. Le graphe de 1,2-inclusion de \mathcal{S}_6 est une étoile. Il n'est pas difficile de constater que $\text{sVC-dim}(\mathcal{S}_6) = m$. D'autre part, pour la famille plantée $\mathcal{S}'_6 := \mathcal{S}_6 \Delta \{e_1, x_1\} = \{\emptyset\} \cup \{\{e_i, x_i\} : 1 \leq i \leq m\}$, il est rapide de vérifier que $\text{cVC-dim}(\mathcal{S}'_6) = 2$, montrant que $\text{cVC-dim}^*(\mathcal{S}_6) = 2$ et que $\text{sVC-dim}^*(\mathcal{S}_6) = m$. Ainsi, $\text{sVC-dim}^*(\mathcal{S})$ peut être arbitrairement supérieure à $\text{cVC-dim}^*(\mathcal{S})$.

\mathcal{S}	VC-dim	cVC-dim	cVC-dim*	dégénérescence	densité	2VC-dim	sVC-dim
\mathcal{S}_0	1	–	m	$m - 1$	$\frac{m-1}{2}$	0	1
\mathcal{S}_1	1	m	m	$m - 1$	$\frac{m-1}{2}$	2	m
\mathcal{S}_2	2	3	3	3	$\frac{3}{2}$	3	3
\mathcal{S}_3	2	$k + 2$	$k + 2$	k	$\frac{k}{2} + o(1)$	2	$k + 1$
$\mathcal{S}_3 \Delta X$	2	$(m - 1)k + 1$	$k + 2$	k	$\frac{k}{2} + o(1)$	2	1
\mathcal{S}_4	$m - 1$	m	m	$\binom{m}{2}$	$\frac{1}{2} \binom{m}{2}$	m	m

TABLE 6.1 – Calculs de quelques paramètres des familles paires plantées $\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_4$ définies dans les exemples 3–6 : la VC-dimension, les deux clique-VC-dimensions, la 2VC-dimension, la dégénérescence et la densité

Il semble donc naturel de se demander s'il est possible, dans le théorème 19, de remplacer $\text{cVC-dim}^*(\mathcal{S})$ par $\text{sVC-dim}^*(\mathcal{S})$. Toutefois, nous ne sommes pas parvenus à répondre à la question suivante :

Question 1. Est-il vrai que, pour toute famille d'ensembles (paire) \mathcal{S} de graphe de 1,2-inclusion $G_{1,2}(\mathcal{S}) = (V, E)$ et de star-VC-dimension $d := \text{sVC-dim}^*(\mathcal{S})$, $\frac{|E|}{|V|} = O(d^2)$?

Une des difficultés principales ici est qu'une opération de d-décalage peut augmenter la star-VC-dimension (autrement dit, le lemme 21 n'est plus vérifié). La différence entre s-pulvérisation et c-pulvérisation est qu'un 2-ensemble $\{e, y\}$ avec $y \in Y$ peut être s-pulvérisé seulement par un ensemble $S \in \mathcal{S}_{ij}$ appartenant à la fibre $F(\{e, y\})$ (c'est la condition $Y \cap S = \{e, y\}$), alors que cette paire $\{e, y\}$ peut être c-pulvérisée par n'importe quel ensemble l'incluant (c'est la condition $\{e, y\} \subseteq S$). Lorsqu'une opération de d-décalage φ_{ij} est effectuée respectivement à une paire $\{e_i, e_j\}$ telle que $\{e_i, e_j\} \cap \{e, y\} = \emptyset$, un ensemble $S \in \mathcal{S}$ peut être envoyé sur un ensemble $\varphi_{ij}(S)$ appartenant à la fibre $F(\{e, y\})$. Si $\varphi_{ij}(S)$ est utilisé pour c-pulvériser le 2-ensemble $\{e, y\}$ par $\varphi_{ij}(\mathcal{S})$, alors S peut être utilisé pour c-pulvériser $\{e, y\}$ par \mathcal{S} (c'est la preuve du lemme 21). Cependant, cela n'est pas vrai pour l'opération de s-pulvérisation car S pourrait ne pas appartenir à $F(\{e, y\})$ au départ. Cela pose donc problème pour une approche par décalage comme nous le faisons dans ce document.

Conclusion

Dans le chapitre 5, nous avons adapté la preuve par induction de HAUSSLER, LITTLESTONE et WARMUTH [100] afin de montrer que, si $G = (V, E)$ est un sous-graphe d'un produit cartésien Γ de m graphes G_1, \dots, G_m excluant un mineur H , alors $\frac{|E|}{|V|} \leq \mu(H) \cdot \text{VC-dim}^*(G)$, où $\mu(H)$ est une constante dépendant de H et où $\text{VC-dim}^*(G)$ est une notion de VC-dimension généralisée aux produits cartésiens de graphes. Notre induction est essentiellement basée sur les lemmes 12, 13 et 14 qui généralisent le lemme 4 pour les sous-graphes d'hypercubes.

Lorsque nous définissons la VC-dimension pour les produits cartésiens de graphes, nous introduisons aussi une notion que nous nommons VC-densité. Cette notion semble assez naturelle pour étudier la densité d'un sous-graphe G d'un produit cartésien Γ . Intuitivement, elle permet de prendre en compte à la fois la VC-dimension de G et la densité maximale des facteurs de Γ (voir le lemme 10). Nous faisons alors la conjecture suivante :

Conjecture (1). *Soit $G = (V, E)$ un sous-graphe d'un produit cartésien de graphes connexes. Alors $\frac{|E|}{|V|} \leq \text{VC-dens}^*(G)$.*

Malheureusement, nous ne sommes pas parvenus à utiliser convenablement cette notion de VC-densité. Pour prouver la conjecture 1 avec notre méthode, il serait nécessaire de montrer une inégalité plus serrée que l'inégalité (2) du lemme 13. Une amélioration de cette inégalité semble possible. En effet, l'opération générant G_c^{uv} contracte l'un des facteurs en un seul sommet. Lorsque nous étudions sa VC-dimension, il n'est pas surprenant qu'elle soit inférieure au moins de 1 à celle de G . Pour ce qui est de sa VC-densité, nous pourrions cependant espérer retirer une quantité de l'ordre de la densité du facteur.

Dans le chapitre 6, nous nous sommes inspirés de l'opération de décalage de HAUSSLER [99] pour montrer que, si \mathcal{S} est une famille d'ensembles paire et que $G_{1,2}(\mathcal{S}) = (V, E)$ désigne son graphe de 1,2-inclusion, alors l'inégalité $\frac{|E|}{|V|} \leq \binom{d}{2}$ est vérifiée, où $d := \text{cVC-dim}^*(\mathcal{S})$ dénote la clique-VC-dimension de \mathcal{S} . Les ingrédients essentiels à notre preuve sont présents dans le lemme 20 (affirmant que les d-décalages préservent le nombre de sommets et ne diminuent pas le nombre d'arêtes), le lemme 21 (montrant que les d-décalages n'augmentent pas la clique-VC-dimension) et le lemme 24 (bornant la densité des bouquets de demi-cubes résultants de séquences complètes de d-décalages). Chacun de ces

résultats est établi pour des familles d'ensembles paires. Les lemmes 20 et 24 ne sont pas vraiment sensibles au choix de la définition de clique-VC-dimension. Elles nécessitent néanmoins l'utilisation de la définition des graphes de 1,2-inclusion en tant que sous-graphes de demi-cubes. Le lemme 21, en revanche, dépend énormément de la définition de la clique-VC-dimension.

Une autre question pour laquelle nous n'avons trouvé ni preuve ni contre-exemple est la suivante :

Question 2. Soit \mathcal{S} une famille d'ensembles de VC-dimension (classique) d . Soit $G := G_{1,2}(\mathcal{S}) = (V, E)$ le graphe de 1,2-inclusion de \mathcal{S} , et soit ω la taille maximale d'une clique de G . Est-il vrai qu'alors $\frac{|E|}{|V|} = O(d \cdot \omega)$?

En effet, nous savons que les hypercubes et les cliques sont deux sous-familles des graphes de 1,2-inclusion. D'une part, seule la VC-dimension joue sur la densité d'un hypercube (puisque les cliques sont de taille 2). D'autre part, c'est évidemment la taille de la clique qui influe sur la densité de celle-ci. Cela explique la nécessité des deux paramètres (VC-dimension et taille de clique) dans la formulation de la question 2. Une des difficultés pour répondre à cette question est qu'une approche par décalage ne semble pas adaptée. En effet, cette opération peut augmenter la taille des cliques du graphe de 1,2-inclusion.

Une approche alternative à la problématique posée par la question 2 (ou même pour la question 1 de la section 6.6) consiste à adapter la preuve originale du théorème 8 de HAUSSLER, LITTLESTONE et WARMUTH [100], plutôt que celle par décalage de HAUSSLER [99]. Malheureusement, comme dans le cas des décalages, la clique maximale de $G_{1,2}(\mathcal{S}_e)$ peut être strictement plus grande que celle de $G_{1,2}(\mathcal{S})$. De ce fait, l'inégalité $|E(G)| \leq |E(G_1(\mathcal{S}_e))| + |E(G_1(\mathcal{S}^e))| + |\mathcal{S}^e|$ n'est pas toujours vérifiée si l'on ne considère pas le graphe de 1-inclusion mais celui de 1,2-inclusion (rappelons que les notations \mathcal{S}_e et \mathcal{S}^e sont introduites dans la section 2.4.2).

Troisième partie

Schémas d'étiquetage de distance et de routage

Introduction

Les représentations classiques de graphes/réseaux sont, pour la plupart, des représentations globales. Si l'on veut récupérer une quelconque information sur le graphe, nous devons accéder à une structure de donnée le représentant dans sa totalité, même si la requête porte sur un nombre limité de nœuds. Avec des réseaux qui, de nos jours, sont de plus en plus grands, utiliser des représentations plus locales peut s'avérer nécessaire. En effet, certains réseaux sont trop grands pour supporter une représentation globale, et dans ces cas, l'information doit être distribuée. La notion de schéma d'étiquetage permettant de répondre de manière distribuée à des requêtes de distance ou de routage, essentielles dans les réseaux, a donc été introduite pour répondre à ce besoin [39, 85, 106, 131, 132]. Dans cette partie, nous nous intéressons à des schémas de distances et de routage pour deux sous-familles des graphes faiblement modulaires : les graphes médians sans cubes et pontés sans K_4 .

Dans le chapitre 7, nous concevons des schémas d'étiquetage de distance et de routage pour la classe des graphes médians sans cube (c.-à-d., 3-cube), avec n sommets. Nos schémas utilisent des étiquettes de $O(\log^3 n)$ bits pour des requêtes s'effectuant en temps constant $O(1)$.

Dans le chapitre 8, nous présentons un schéma de distance approché pour la famille des graphes pontés sans K_4 , avec n sommets. Ce schéma utilise des étiquettes de $O(\log^3 n)$ bits. La fonction de décodage calcule, en temps constant $O(1)$, à partir de deux étiquettes, la distance entre les sommets qu'elles encodent en faisant une erreur multiplicative d'au plus 4.

Avant de parler de nos schémas pour les graphes médians ou pontés, nous pouvons rappeler un certain nombre de faits déjà connus. Tout d'abord, comme nous l'avons expliqué en chapitre 2, un graphe médian G avec n sommets possède au plus $n \cdot d$ arêtes, où d est la plus grande dimension d'un cube sous-graphe de G . Nous pouvons en déduire qu'ils admettent un schéma d'adjacence avec étiquettes de taille $O(d \log n)$ (voir section 3.2.1). Cela conduit au fait que les graphes médians sans cubes admettent un schéma d'adjacence avec des étiquettes de longueur logarithmiques. En ce qui concerne les graphes pontés, ils sont démontables (par des sommets simpliciaux). Ainsi, leur dégénérescence et leur densité dépendent de la taille de leur plus grande clique. Si cette dernière est fixée (comme dans les graphes pontés sans K_4), alors les étiquettes attribuées

par un schéma d'adjacence sont aussi de longueur logarithmique.

Concernant des schémas de distance et routage compacts, il en existe déjà pour certaines sous-classes des graphes médians sans cube et des graphes pontés sans K_4 . Par exemple, la classe des graphes planaires, avec n sommets, et dans lesquels chaque sommet interne a degré au moins 4 et chaque face est de longueur au moins 4 (c.-à-d., les $(4, 4)$ -graphes planaires), admet des schémas de distance et de routage avec étiquettes de longueur $O(\log^2 n)$. Les graphes médians sans cubes planaires sont de tels graphes. De même, la classe des graphes planaires, avec n sommets, et dans lesquels les sommets internes ont degré au moins 6 et les faces sont de longueur au moins 3 (c.-à-d., les $(3, 6)$ -graphes planaires) admet aussi de tels schémas. Ils contiennent, quant à eux, les graphes pontés planaires sans K_4 . Ces deux résultats découlent en fait d'un même résultat plus général portant sur les graphes planaires de courbure non-positive, montré par CHEPOI, DRAGAN et VAXÈS [53].

Une autre sous-classe des médians sans cube admettant un schéma de distance avec étiquettes de $O(\log^2 n)$ bits est celles des *doubles arbres partiels*, ce sont les graphes admettant un plongement isométrique dans un produit cartésien de deux arbres. Ces graphes sont aussi caractérisés par BANDELT, CHEPOI et EPPSTEIN [24] comme les graphes médians sans cube dans lesquels le "link" de chaque sommet est biparti. Cette borne supérieure de $O(\log^2 n)$ dans les doubles arbres partiels se déduit en fait directement de leur plongement isométrique dans le produit cartésien de deux arbres.

Chapitre 7

Graphes médians sans cubes

Sommaire

7.1	Résultats principaux	116
7.2	L'idée du schéma	118
7.3	Étoiles dans les graphes médians	119
7.3.1	Schémas de distance et de routage pour les étoiles	120
7.4	Fibres des graphes médians	121
7.5	Fibres dans les graphes médians sans cube	125
7.5.1	Classification des fibres	126
7.5.2	La bordure totale des fibres est quasi-portée	126
7.6	Classification des paires de sommets et plus courts chemins	130
7.7	Schéma de distance pour les graphes médians sans cube	133
7.7.1	Encodage	134
7.7.2	Requêtes de distance	136
7.7.3	Correction et complexité	139
7.8	Schéma de routage pour les graphes médians sans cube	141
7.8.1	L'idée	141
7.8.2	Encodage	142
7.8.3	Requêtes de routage	144

Dans ce chapitre, nous montrons que la classe des graphes médians sans cube sur n sommets admet des schémas de distance et de routage utilisant des étiquettes de $O(\log^3 n)$ bits. Dans les deux cas, ces étiquettes sont attribuées en temps polynomial et décodées en temps constant.

7.1 Résultats principaux

Un graphe médian G est dit *sans cube* s'il est de dimension 2 (c.-à-d., s'il ne possède pas de 3-cube en tant que sous-graphe isométrique). Deux illustrations

de graphes médians sans cube sont données dans la figure 7.1. Bien que les graphes médians sans cube ne correspondent qu'au 1-squelette des complexes cubiques CAT(0) 2-dimensionnels, leur structure combinatoire est tout de même assez complexe. Entre autres choses, toute clique peut être obtenue comme mineur d'un graphe médian sans cubes. En particulier, ces derniers ne sont pas nécessairement planaires. Par exemple, le produit cartésien $K_{1,n} \square K_{1,m}$ des étoiles $K_{1,n}$ et $K_{1,m}$ pour $n, m \geq 5$ est un graphe médian sans cube non planaire.

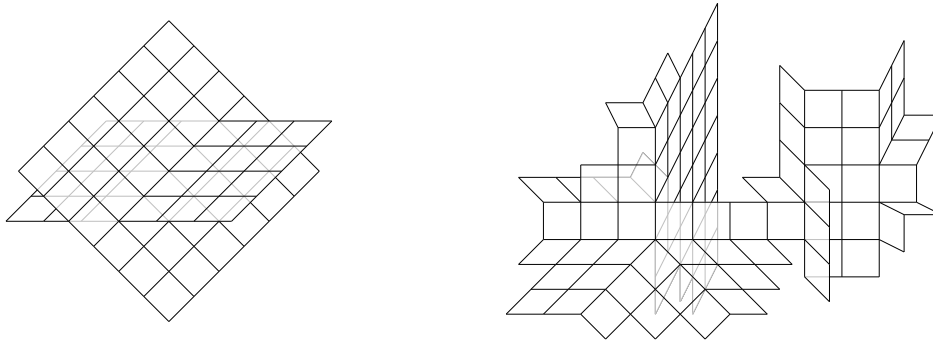


FIGURE 7.1 – Deux graphes médians sans cube. Celui de gauche fera office d'exemple de travail.

Le chapitre est organisé comme suit. Dans la section 7.2, nous cherchons à donner un peu d'intuition pour la suite en décrivant “avec les mains” l'idée de nos schémas. Dans les section 7.3 et 7.4, nous introduisons (respectivement) les notions d'étoiles et de fibres dans un graphe médian qui sont les ingrédients principaux de la partition à la base de notre schéma. Nous énonçons aussi une partie de leurs propriétés valables dans les graphes médians quelconques. Dans la section 7.5, nous continuons avec les propriétés essentielles de ces deux objets dans le cas des graphes médians sans cube. Nous donnons ensuite, dans la section 7.6, une classification des paires de sommets dans les graphes médians sans cubes. Cette classification nous permet de décrire des plus courts chemins reliant les paires de sommets en fonction de leur classe. La section 7.7 fait l'objet de la description de notre schéma de distance pour les graphes médians sans cube, ainsi que de la preuve du théorème suivant :

Théorème 20. *La classe \mathcal{G} des graphes médians sans cube avec n sommets admet un schéma d'étiquetage de distance utilisant des étiquettes de $O(\log^3 n)$ bits, attribuées en temps $O(n^2 \log n)$. Étant données deux telles étiquettes, la distance exacte entre les deux sommets qu'elles encodent peut être calculée en temps constant $O(1)$.*

Enfin, dans la section 7.8, nous adaptons le schéma de distance pour obtenir un schéma de routage pour les graphes médians sans cubes. Nous obtenons alors le théorème qui suit.

Théorème 21. *La classe \mathcal{G} des graphes médians sans cube avec n sommets admet un schéma d'étiquetage de routage utilisant des étiquettes de $O(\log^3 n)$ bits, attribuées en temps $O(n^2 \log n)$. Étant données deux étiquettes, attribuées à deux sommets u et v d'un graphe $G \in \mathcal{G}$, un port à prendre depuis u pour atteindre v via un plus court chemin est calculé en temps constant $O(1)$.*

7.2 L'idée du schéma

Nous commençons par décrire succinctement l'idée du schéma pour le cas des distances. Le cas du routage est assez similaire, ses différences avec le schéma de distance seront traitées en section 7.8.1.

Soit $G = (V, E)$ un graphe médian sans cube avec n sommets. L'algorithme commence par rechercher un sommet médian m de G et calcule l'étoile $\text{St}(m)$ de m (c.-à-d., l'union des arêtes et carrés incidents à m). Nous montrons que l'étoile $\text{St}(m)$ est portée et pouvons construire une partition de V en fibres $F(x)$: pour $x \in \text{St}(m)$, $F(x)$ consiste en l'ensemble de sommets de V de porte x sur $\text{St}(m)$. Puisque m est un sommet médian de G , chaque fibre contient au plus la moitié des sommets de G . Nous montrons, de plus, que ces fibres sont portées. Nous distinguons deux types de fibres $F(x)$ selon la distance entre x et m : les panneaux pour une distance 1, les cônes pour une distance 2. Nous montrerons qu'un cône est "voisin" d'exactly deux panneaux, mais que les panneaux peuvent être voisins d'un nombre arbitraires de cônes. Étant donné deux sommets arbitraires u et v de G , nous montrons que $d_G(u, v) = d_G(u, m) + d_G(m, v)$ lorsque les fibres contenant u et v sont distinctes et suffisamment "éloignées". Dans le cas plus problématique de fibres "trop proches", nous distinguerons deux autres cas : un panneau 1-voisin d'un cône, et un cône 2-voisin d'un autre cône. Lorsque $d_G(u, v) = d_G(u, m) + d_G(m, v)$, la distance entre u et v peut être récupérée en stockant essentiellement $d_G(u, m)$ dans l'étiquette de u et $d_G(v, m)$ dans celle de v . Si u et v appartiennent à une même fibre $F(x)$, cette dernière sera récursivement partitionnée, les informations pour calculer $d_G(u, v)$ seront donc récupérées à une étape future du partitionnement. Pour les deux cas restant, nous verrons que $d_G(u, v)$ pourra être obtenue en stockant des informations (courtes) relatives seulement à deux "projections" de u et v sur des fibres voisines à la leur, cela pour essentiellement deux raisons. D'une part, les cônes ne sont voisins que de deux panneaux. D'autre part (et c'est la contribution principale du chapitre), il s'avère que, pour chaque panneau $F(x)$, l'union des projections des cônes voisins vers $F(x)$ est incluse dans un arbre isométrique de G , et les sommets de $F(x)$ n'admettent qu'au plus deux projections sur cet arbre. Il suffira donc de conserver ces projections pour chaque $u \in F(x)$ ainsi qu'un étiquetage de distance dans l'arbre isométrique des projections. Il en résultera que seule-

ment $O(\log^2 |V|)$ bits sont à ajouter aux étiquettes à chaque étape récursive. La profondeur de la récursion étant bornée par $O(\log |V|)$, les étiquettes finales seront de longueur $O(\log^3 |V|)$. La distance $d_G(u, v)$ pourra être calculée en trouvant dans les étiquettes le premier médian ayant “séparé” u et v (et en utilisant les informations des deux projections à cette étape).

7.3 Étoiles dans les graphes médians

L'étoile $\text{St}(z)$ d'un sommet z dans un graphe médian $G = (V, E)$ consiste en l'union de tous les hypercubes de G contenant z (c.-à-d., c'est le plus grand bouquet de cubes centré en z).

D'un point de vue combinatoire, les étoiles peuvent avoir une structure assez arbitraire dans les graphes médians. En effet, BANDELT et DE VEL [25] montrent que les étoiles dans les graphes médians sont en bijection avec l'ensemble des graphes. Cette bijection est donnée via la construction du *graphe simplexe* $\sigma(H)$ d'un graphe quelconque H . $\sigma(H)$ est le graphe dont l'ensemble des sommets correspond à l'ensemble des cliques de H (c.-à-d., ensemble vide, sommets, arêtes, triangles, etc.) et dans lequel deux cliques sont reliées si elles diffèrent sur un unique sommet (voir la figure 7.2 pour une illustration). BANDELT et DE VEL montrent alors que $\sigma(H)$ est un graphe médian. De plus, il est facile de remarquer que l'étoile du sommet v_\emptyset correspondant à l'ensemble vide dans $\sigma(H)$ coïncide avec $\sigma(H)$ tout entier. De manière équivalente, toute étoile $\text{St}(z)$ d'un sommet z d'un graphe médian peut être réalisée comme le graphe simplexe du graphe H ayant $N(z)$ pour ensemble de sommets et dans lequel deux voisins u' et u'' de z sont adjacents (dans H) si et seulement si z, u' et u'' appartiennent à un même carré (de G).

Considérons maintenant les étoiles $\text{St}(z)$ des graphes médians d'un point de vue métrique.

Lemme 25. *Pour tout sommet z d'un graphe médian $G = (V, E)$, $\text{St}(z)$ est un sous-graphe porté de G .*

Démonstration. Par le lemme 6, il suffit de montrer que $\text{St}(z)$ est localement convexe. Soit $x, y \in \text{St}(z)$ deux sommets à distance 2 l'un de l'autre, et soit v un voisin commun de x et de y . Alors $Q_x := I(x, z)$ et $Q_y := I(y, z)$ sont deux cubes de $\text{St}(z)$. Si $v \in Q_x \cup Q_y$, il ne constitue pas un contre-exemple à la convexité locale de $\text{St}(z)$. Supposons donc que $v \notin Q_x \cup Q_y$. Cela implique que $x, y \in I(z, v)$ (c.-à-d., nous pouvons supposer que $d_G(x, z) = d_G(y, z) = k$ et $d_G(z, v) = k + 1$). Par la condition du quadrangle (sur z, x, y et v), il doit exister un sommet u tel que $d_G(u, z) = k - 1$ et $u \sim x, y$. Forcément, $I(u, z)$ est un $(k - 1)$ -cube Q_u inclut

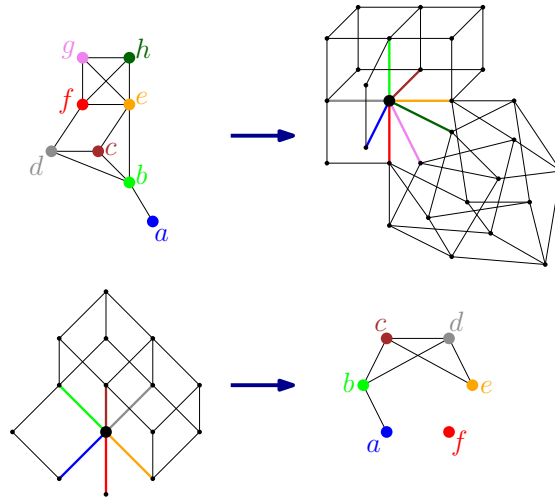


FIGURE 7.2 – Sur la première ligne, un exemple d'un graphe H (à gauche) et du graphe simplexe correspondant $\sigma(H)$ (à droite). Sur la ligne suivante, la construction d'un graphe à partir d'un bouquet de cubes $\text{St}(z)$.

dans les k -cubes Q_x et Q_y . Par conséquent, z possède un voisin x' tel que $I(x, x')$ soit un $(k - 1)$ -cube $Q_{x'}$ disjoint de Q_u et qui, pris avec Q_u , donne Q_x . De façon analogue, z possède un voisin y' tel que $I(y, y')$ soit un $(k - 1)$ -cube $Q_{y'}$ disjoint de Q_u et qui, pris avec Q_u , donne Q_y . En appliquant la condition du quadrangle (aux sommets x', y', z et v), nous obtenons l'existence d'un sommet $v' \sim x', y'$ à distance $k - 1$ de v . Nous pouvons alors montrer que $I(v, v')$ induit un $(k - 1)$ -cube Q_v par application successive de la condition du quadrangle entre v et des sommets de $Q_{x'}$, $Q_{y'}$ et Q_u . Ainsi, $Q_{x'}$, $Q_{y'}$, Q_v et Q_u forment un cube $(k + 1)$ dimensionnel $I(v, z)$ et v doit donc appartenir à $\text{St}(z)$ (la figure 7.4(a) illustre les notations utilisées dans cette preuve). \square

7.3.1 Schémas de distance et de routage pour les étoiles

Nous présentons ici un schéma de distance (et de routage) pour les étoiles $\text{St}(z)$ des graphes médians G . Ce schéma est basé sur le fait qu'un graphe médian admet toujours un plongement isométrique dans un hypercube et que $\text{St}(z)$ est un sous-graphe porté et est donc un sous-graphe isométrique du graphe médian G . Nous pouvons donc affirmer que $\text{St}(z)$ admet un plongement isométrique dans un hypercube. Soit $\varphi : \text{St}(z) \rightarrow 2^{\{1, \dots, d\}}$ un plongement isométrique de $\text{St}(z)$ dans Q_d tel que $\varphi(z) = \emptyset$. Alors, pour chaque sommet x de $\text{St}(z)$, $\varphi(x)$ consiste en un ensemble de cardinal égal à la dimension du cube $I(x, z)$. $\varphi(x)$ a donc cardinal au plus $\log |\text{St}(z)|$. Pour toute paire de sommets (x, y) de $\text{St}(z)$, $d_{\text{St}(z)}(x, y) = |\varphi(x) \Delta \varphi(y)|$.

En utilisant ce plongement isométrique φ , nous pouvons décrire un encodage simple $\text{Enc_Etoile_Med}(\text{St}(z))$ des sommets de $\text{St}(z)$ pouvant être utilisé pour répondre à la fois à des requêtes de routage et de distance. Pour un sommet $x \in \text{St}(z)$, soit $L_{\text{St}(z)}(x) = \varphi(x)$. Alors $\text{Enc_Etoile_Med}(\text{St}(z))$ donne à z l'étiquette \emptyset et à chaque voisin de z une étiquette unique dans $\{1, \dots, \text{deg}(z)\}$. Pour chaque x à distance k de z , $I(x, z)$ contient exactement k voisins de z , et les étiquettes de ces voisins définissent complètement $\varphi(x)$ et $L_{\text{St}(z)}(x)$ (rappel de la section 1.2.3). Donner un identifiant unique à chaque voisin de z requiert $\lceil \log(\text{deg}(z)) \rceil$ bits. Dans le pire des cas, $\text{Enc_Etoile_Med}(\text{St}(z))$ distribue donc des étiquettes de longueur $O(\text{deg}(z) \cdot \log(\text{deg}(z)))$. Si la dimension de $\text{St}(z)$ est une constante fixée, alors $\text{Enc_Etoile_Med}(\text{St}(z))$ donne des étiquettes de longueur logarithmique.

Par souci de simplicité, nous supposons que, pour un sommet x étiqueté par X et un sommet x' étiqueté par $X' = X \setminus \{i\}$, $i \in X$, nous avons $\text{port}(x, x') = \text{port}(x', x) = i$. Étant donné que φ est un plongement isométrique, il n'est pas difficile de voir que, pour n'importe quels deux sommets x et y encodés par les ensembles $X := L_{\text{St}(z)}(x) = \varphi(x)$ et $Y := L_{\text{St}(z)}(y) = \varphi(y)$, la distance $d_{\text{St}(z)}(x, y)$ entre x et y vaut $|X \Delta Y|$ (voir la figure 7.3). C'est donc exactement la valeur retournée par $\text{Dist_Etoile}(X, Y)$. La décision pour le routage découle de la même propriété. Supposons que $|X| \leq |Y|$. Si $X \subseteq Y$, $\text{Rout_Etoile}(X, Y)$ retourne le port vers n'importe quel sommet étiqueté par $X \cup \{i\}$ avec $i \in Y \setminus X$ (disons le i minimal). Si $X \not\subseteq Y$, alors $\text{Rout_Etoile}(X, Y)$ retourne le port vers n'importe quel sommet étiqueté par $X \setminus \{i\}$ pour $i \in X \setminus Y$ (disons, à nouveau, le i minimal).

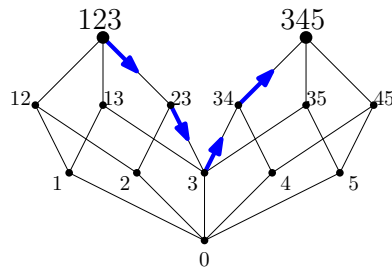


FIGURE 7.3 – Exemple de l'étoile d'un sommet plongée dans un 5-cube, et de la décision de routage pour partir du sommet 123 en direction du sommet 345. La distance entre eux est $|\{1, 2, 4, 5\}| = 4$.

7.4 Fibres des graphes médians

Nous continuons avec la définition des fibres dans un graphe médian $G = (V, E)$, ainsi qu'avec l'énoncé de certaines de leurs propriétés.

Pour un sommet x d'un sous-graphe porté H de G , l'ensemble (ou le sous-graphe induit par cet ensemble) $F(x) := \{v \in V : x \text{ est la porte de } v \text{ dans } H\}$ est appelé *fibres* de x relativement à H . De par leur définition, il est facile (et néanmoins essentiel) d'observer la propriété suivante des fibres de H :

Proposition 12. *L'ensemble $\mathcal{F} := \{F(x) : x \in V(H)\}$ des fibres des sommets de H constitue une partition de G .*

La propriété qui suit est bien connue dans les graphes *fibres-complémentés* ("fiber-complemented") étudiés par CHASTAND [45]. Les graphes médians, étant fibres-complémentés, ils héritent de cette propriété.

Lemme 26. *Pour tout sous-graphe porté H d'un graphe médian G , les fibres $F(x)$ avec $x \in V(H)$ sont portées.*

Démonstration. Toute fibre $F(x)$ induit un sous-graphe connexe de G . Il suffit donc de démontrer que chaque $F(x)$ est localement convexe. Pour cela, considérons deux sommets u et v de $F(x)$ avec $d_G(u, v) = 2$. Soit w un voisin commun de u et v que nous supposons, par l'absurde, ne pas appartenir à $F(x)$ (c.-à-d., $w \in F(y)$ pour un certain $y \in V(H)$, $y \neq x$). Alors $x \in I(u, y) \cap I(v, y)$ et $y \in I(w, x)$. Cela implique en particulier que $x \sim y$, que $u, v \in I(w, x)$ et que $w \in I(u, y) \cap I(v, y)$. Par la condition du quadrangle (sur u, v, w et x), il existe un sommet $x' \sim u, v$ d'un cran plus proche de x que u et v . Donc $w, x' \in I(u, y)$, et par la condition du quadrangle (appliquée à u, x', w et y), il existe un sommet $y' \sim x', z$ d'un cran plus proche de y que x' et w . Mais alors les sommets u, v, w, x' et y' induisent un $K_{2,3}$, qui est un sous-graphe interdit dans un graphe médian (la figure 7.4(b) illustre cette preuve). \square

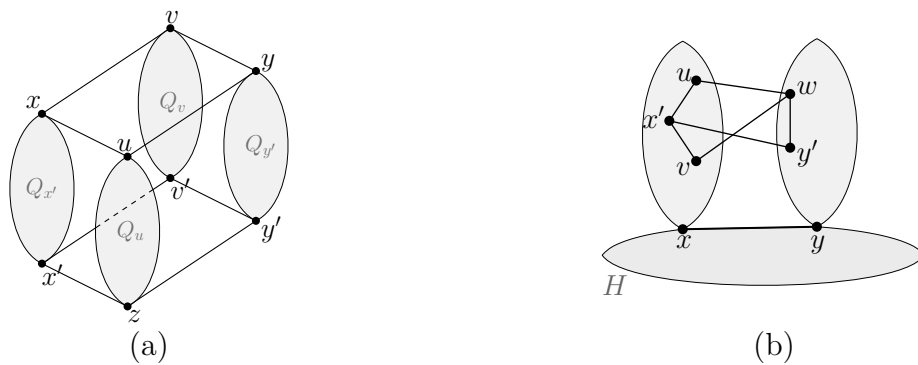


FIGURE 7.4 – La figure (a) illustre la preuve du lemme 25 ; la (b) illustre la preuve du lemme 26.

Le lemme 26 possède deux corollaires. Le premier découle de l'utilisation de ce résultat avec le lemme 25. Nous obtenons alors le résultat suivant

Corollaire 7. *Pour tout sommet z d'un graphe médian G , les fibres du bouquet de cube $\text{St}(z)$ sont portées.*

Puisque les arêtes d'un graphe médian G sont portées, en utilisant le lemme 26 pour les arêtes de G , nous obtenons le second corollaire :

Corollaire 8. *Pour toute arête uv d'un graphe médian G , les demi-espaces $W(u, v)$ et $W(v, u)$ sont portés.*

Le fait que les demi-espaces d'un graphe médian soient convexes a été montré pour la première fois par MULDER [118]. Par la même occasion, l'auteur a établi que la frontière de chaque demi-espace $W(u, v)$ (c.-à-d., l'ensemble $\partial W(u, v) = \{z' \in W(u, v) : \exists z'' \in W(v, u), z'' \sim z'\}$) est aussi convexe. Dans la suite, nous montrerons que cette propriété qu'ont les frontières à être portées est vérifiée pour tout sous-graphe porté d'un graphe médian.

Soit H un sous-graphe porté d'un graphe médian $G = (V, E)$ et soit $\mathcal{F}(H) := \{F(x) : x \in V(H)\}$ la partition de V en les fibres de H . Deux fibres $F(x)$ et $F(y)$ sont dites voisines (et notées $F(x) \sim F(y)$) s'il existe une arête dans G dont l'une des extrémités est dans $F(x)$ et l'autre dans $F(y)$. Si $F(x)$ et $F(y)$ sont deux fibres voisines de H , alors nous dénotons par $\partial_{F(y)}F(x)$ l'ensemble des sommets de $F(x)$ reliés par une arête à un sommet de $F(y)$ (c.-à-d., $\partial_{F(y)}F(x) := \{u \in F(x) : \exists v \in F(y) \text{ t.q. } uv \in E\}$). Cet ensemble est appelé *frontière de $F(x)$ vers $F(y)$* .

Lemme 27. *Soit H un sous-graphe porté d'un graphe médian $G = (V, E)$. Deux fibres $F(x)$ et $F(y)$ de H sont voisines si et seulement si $x \sim y$. De plus, si $F(x) \sim F(y)$, alors la frontière $\partial_{F(y)}F(x)$ de $F(x)$ vers $F(y)$ induit un sous-graphe porté de G de dimension $\leq \dim(G) - 1$.*

Démonstration. Si $x \sim y$, alors $F(x)$ et $F(y)$ sont forcément voisines. Réciproquement, $x' \in F(x)$ et $y' \in F(y)$ tels que $x'y' \in E$ alors, puisque $F(x)$ et $F(y)$ sont convexes et puisque G est biparti, nous avons $x' \in I(y', x)$ et $y' \in I(x', y)$. Puisque $x' \in F(x)$, $y' \in F(y)$ et puisque H est porté, nous déduisons que $x \in I(x', y)$ et $y \in I(y', x)$. Nous obtenons alors que $d_G(x, x') = d_G(y, y')$. Nous pouvons donc conclure que $d_G(x, y) = d_G(x', y') = 1$, ce qui établit la première assertion.

Afin de montrer la seconde affirmation, considérons deux fibres voisines $F(x)$ et $F(y)$ et montrons que $\partial_{F(y)}F(x)$ est porté. Commençons par montrer, par induction sur $k := d_G(x, x')$, que $I(x', x) \subseteq \partial_{F(y)}F(x)$ pour tout sommet x' de $\partial_{F(y)}F(x)$. Pour se faire, il suffit de montrer que tout voisin x'' de x' dans $I(x', x)$ appartient à $\partial_{F(y)}F(x)$. Soit y' le voisin de x' dans $\partial_{F(y)}F(y)$. Alors $x'', y' \in I(x', y)$, $x'', y' \sim x'$ et $d_G(x', y) = k + 1$. Par la condition du quadrangle (sur x'', x', y' et y), nous déduisons qu'il existe un sommet $y'' \sim y', x''$ à distance $k - 1$ de y .

Puisque $y'' \in I(y', y) \subseteq F(y)$, nous concluons que $x'' \in \partial_{F(y)}F(x)$. Par conséquent, $I(x', x) \subseteq \partial_{F(y)}F(x)$ et le sous-graphe induit par $\partial_{F(y)}F(x)$ est connexe.

D'après le lemme 6, il nous reste à montrer que $\partial_{F(y)}F(x)$ est localement convexe. Soient $x', x'' \in \partial_{F(y)}F(x)$ deux sommets à distance deux l'un de l'autre et soit $u \sim x', x''$. Puisque $F(x)$ est convexe, $u \in F(x)$. Soit y' et y'' les voisins respectifs de x' et x'' dans $F(y)$. Soit v la porte de u sur $F(y)$ (qui doit exister puisque $F(y)$ est porté par le lemme 26). Comme $d_G(u, y') = d_G(u, y'') = 2$ (car G est biparti) et $v \in I(u, y') \cap I(u, y'')$, v doit être adjacent à u, y' et y'' . Ainsi, $v \in I(y', y'') \subseteq F(y)$, conduisant à $u \in \partial_{F(y)}F(x)$ et achevant la preuve du fait que $\partial_{F(y)}F(x)$ soit porté.

Si $\dim(\partial_{F(y)}F(x)) = \dim(G) =: d$, alors $\partial_{F(y)}F(x)$ contient un cube d -dimensionnel Q' . Mais alors les voisins des sommets de Q' dans $\partial_{F(x)}F(y)$ induisent aussi un d -cube Q'' , et Q' et Q'' forment alors un cube $(d + 1)$ -dimensionnel. Cela contredit l'hypothèse $\dim(G) = d$ et montre que $\dim(\partial_{F(y)}F(x)) \leq d - 1$. \square

Pour un sommet x d'un sous-graphe porté H de G , l'union des frontières $\partial_{F(y)}F(x)$ de la fibre $F(x)$ vers chacune de ses voisines $F(y)$ est appelée la *bordure de $F(x)$* , dénotée par $\partial^*F(x)$. C'est-à-dire,

$$\partial^*F(x) := \bigcup_{y \in N_G(x)} \partial_{F(y)}F(x).$$

Lemme 28. *Soit H un sous-graphe porté d'un graphe médian G de dimension d . Alors la bordure $\partial^*F(x)$ de toute fibre $F(x)$ de H ne contient jamais de cube d -dimensionnel.*

Démonstration. Supposons, par l'absurde, que $\partial^*F(x)$ contiennent un cube d -dimensionnel Q . Comme Q doit être un sous-graphe porté de G , nous pouvons considérer la porte x' de x sur Q ainsi que le sommet de Q le plus éloigné de x que nous noterons x'' (autrement dit, x'' est le sommet opposé à x' dans Q). Supposons que $x'' \in \partial_{F(y)}F(x) \subseteq \partial^*F(x)$. Puisque $\partial_{F(y)}F(x)$ est porté (par le lemme 27), et puisque $Q \subseteq I(x, x'')$, Q est forcément inclus entièrement dans $\partial_{F(y)}F(x)$. Mais alors, de même que dans la preuve du lemme 27, considérer l'union de Q avec son cube "voisin" dans $\partial_{F(x)}F(y)$ crée un cube $(d + 1)$ -dimensionnel, et donc une contradiction. \square

Lemme 29. *Soit H un sous-graphe porté d'un graphe médian G . Alors la bordure $\partial^*F(x)$ de toute fibre $F(x)$ de H est un sous-graphe isométrique de G .*

Démonstration. Soient $u, v \in \partial^*F(x)$, disons $u \in \partial_{F(y)}F(x)$ et $v \in \partial_{F(z)}F(x)$. Soit w le médian du triplet (x, u, v) . Puisque $w \in I(u, x) \subseteq \partial_{F(y)}F(x) \subseteq \partial^*F(x)$, nous déduisons que $I(v, w) \subseteq \partial^*F(x)$. De même, nous pouvons montrer que $I(v, w) \subseteq \partial^*F(x)$. Comme $w \in I(u, v)$ et $I(u, w) \cup I(w, v) \subseteq \partial^*F(x)$, les sommets

u et v peuvent être connectés, dans $\partial^*F(x)$, par un plus court chemin passant par w . \square

Nous concluons cette section avec une propriété supplémentaire des fibres de l'étoile de sommet médian de G .

Lemme 30. *Soit m un sommet médian d'un graphe médian G avec n sommets. Alors toute fibre $F(x)$ du bouquet de cube $\text{St}(m)$ possède au plus $n/2$ sommets.*

Démonstration. Supposons, par l'absurde, que $|F(x)| > n/2$ pour un certain sommet $x \in \text{St}(m)$. Soit u un voisin de m dans $I(x, m)$. Pour tout $v \in F(x)$, $x \in I(v, m)$ et $u \in I(x, m)$, et nous pouvons conclure que $u \in I(v, m)$. Par conséquent $F(x) \subseteq W(u, m)$, et donc $|W(u, m)| > n/2$. Il en résulte que $|W(m, u)| = n - |W(u, m)|$, mais cela contredit le fait que m ait été choisi comme un sommet médian de G . En effet, comme $u \sim m$, il est facile de constater que $M(u) - M(m) = |W(m, u)| - |W(u, m)| < 0$. \square

Malheureusement, la bordure $\partial^*F(x)$ d'une fibre $F(x)$ n'induit pas toujours un graphe médian. De ce fait, même si $\partial^*F(x)$ est un sous-graphe isométrique de G de dimension au plus $\dim(G) - 1$, l'étiquetage ne pourra pas être récursivement appliqué au sous-graphe induit par la bordure. Toutefois, si G est sans cube, alors la bordure des fibres est toujours un sous-arbre isométrique de G , et il est donc possible d'utiliser pour elles des schémas de distance ou de routage pour les arbres (graphes médians 1-dimensionnels). Mais, même dans ce cas, il nous reste certaines bonnes propriétés des bordures à établir. C'est l'objet de la section 7.5.

7.5 Fibres dans les graphes médians sans cube

Dans cette section, nous établissons les propriétés supplémentaires que possèdent les fibres des étoiles, ainsi que leur bordure, dans le cas particulier des graphes médians sans cube $G = (V, E)$. En utilisant ces propriétés, nous pourrions ensuite montrer que, pour n'importe quelle paire de sommets u et v de G , la trichotomie suivante est vérifiée. La distance $d_G(u, v)$ est : soit égale à $d_G(u, m) + d_G(m, v)$ (pour un sommet médian m de G) ; soit égale à la somme des distances de u et v vers deux sommets appropriés u' et v' de $\partial^*F(x)$ plus la distance entre u' et v' , soit calculée grâce à un appel récursif sur la fibre contenant u et v .

7.5.1 Classification des fibres

Les graphes $G = (V, E)$ considérés dans le reste de ce chapitre seront supposés médians et sans cube. L'étoile $\text{St}(z)$ de tout sommet z de G consiste donc maintenant en l'union d'arêtes et de carrés contenant z . De même que les étoiles des sommets dans les graphes médians quelconques sont en bijection avec les graphes quelconques (via la construction du graphe simplexe), les étoiles des sommets dans le cas présent des graphes médians sans cubes sont en bijection avec les graphes sans triangles.

Soit z un sommet quelconque de G et soit $\mathcal{F}_z := \{F(x) : x \in \text{St}(z)\}$ la partition de V en les fibres de $\text{St}(z)$. Nous pouvons distinguer deux types de fibres dans \mathcal{F}_z : une fibre $F(x)$ telle que x soit adjacent à z est appelée un *panneau* ; une fibre $F(x)$ telle que $d_G(x, z) = 2$ est appelée un *cône*. Si $F(x)$ est un panneau, alors l'intervalle $I(x, z)$ est une arête. Si $F(x)$ est un cône, alors l'intervalle $I(x, z)$ est un carré $Q_x := (x, y', z, y'')$. Dans le second cas, puisque y' et y'' sont les seuls voisins de x dans $\text{St}(z)$, alors nous déduisons, de par le lemme 27, que $F(x)$ est voisin des panneaux $F(y')$ et $F(y'')$, et d'aucune autre fibre. Par le même lemme 27, nous pouvons observer qu'un panneau $F(y)$ ne peut être voisin d'aucun autre panneau mais peut, en revanche, être voisin d'un nombre quelconque (linéaire en la taille du graphe) de cônes (c.-à-d., tous les cônes $F(x')$ tels que le carré $Q_{x'}$ contienne l'arête yz). La figure 7.6 donne une illustration de ce fait.

7.5.2 La bordure totale des fibres est quasi-portée

La propriété principale d'une empreinte sur un ensemble $A \subseteq V$ est que, pour tout sommet $z \in A$, il existe un plus court (u, z) -chemin passant par $\Upsilon(u, A)$ (c.-à-d., $I(u, z) \cap \Upsilon(u, A) \neq \emptyset$). De ce fait, si l'ensemble $\Upsilon(u, A)$ possède un nombre constant d'éléments, alors nous pouvons stocker dans l'étiquette de u la distance à chaque élément de $\Upsilon(u, A)$. En utilisant cela, pour tout sommet $z \in A$, nous sommes alors capable de calculer $d_G(u, z)$ comme $\min\{d_G(u, a) + d_G(a, z) : a \in \Upsilon(u, A)\}$. L'objectif principal de cette sous-section va être de montrer que la bordure des fibres d'un graphe médian sans cube possède cette propriété, et que c'est un arbre quasi-porté.

Lemme 31. *La bordure $\partial^*F(x)$ de toute fibre $F(x)$ d'un bouquet de cube $\text{St}(z)$ dans un graphe médian sans cube G est un arbre isométrique dont les branches sont portées.*

Démonstration. Des lemmes 28 et 29, il résulte que $\partial^*F(x)$ est un arbre isométrique enraciné en x . Pour tout sommet $v \in \partial^*F(x)$, il existe une fibre $F(y) \sim F(x)$ de

$\text{St}(z)$ telle que v appartienne à la frontière $\partial_{F(y)}F(x)$ de $F(x)$ vers $F(y)$. De par le lemme 27, $\partial_{F(y)}F(x)$ est un sous-arbre porté de G auquel appartiennent v et x . L'unique chemin $P(v, x)$ entre v et x dans $\partial_{F(y)}F(x)$ est un sous-chemin convexe de $\partial_{F(y)}F(x)$ et donc un sous-chemin convexe dans le graphe G tout entier. Puisque les sous-graphes convexes sont portés (lemme 6), $P(x, v)$ est un chemin porté de G inclus dans $\partial^*F(x)$. \square

Par le lemme 31, les branches de $\partial^*F(x)$ sont portées. En revanche, $\partial^*F(x)$ lui-même n'est pas nécessairement porté. Puisqu'un panneau $F(x)$ peut être voisin d'un nombre arbitraire de cônes, il serait intuitif de penser que l'empreinte $\Upsilon(u, \partial^*F(x))$ d'un sommet u du panneau sur sa bordure puisse aussi être de taille arbitrairement grande. Le lemme 32 suivant montre que ce n'est pas le cas. Plus précisément, il montre que $|\Upsilon(u, \partial^*F(x))| \leq 2$, et ce résultat est fondamental dans l'élaboration des schémas de distance et de routage présentés en sections 7.7 et 7.8. Malheureusement, cette propriété n'est plus vraie lorsque la dimension du graphe médian est strictement supérieure à 2.

Lemme 32. *Soit T un arbre enraciné avec des branches portées dans un graphe médian sans cube $G = (V, E)$. Alors T est quasi-porté.*

Démonstration. Soit r la racine de T . Considérons un sommet $u \in V \setminus V(T)$ et faisons l'hypothèse absurde que $\Upsilon(u, T)$ contienne trois empreintes distinctes x_1, x_2 et x_3 . Comme les branches de T sont portées, aucun des sommets x_1, x_2 et x_3 ne peut appartenir à un même chemin de T entre la racine et l'un de ces trois sommets. En particulier, r doit être différent de x_1, x_2 et x_3 . Supposons, de plus, que parmi tous les arbres enracinés T' dans G , avec des branches portées, et tels que $|\Upsilon(u, T')| \geq 3$, T ait un nombre de sommet minimal. Par ce choix de minimalité (et du fait que chaque sous-arbre de T contenant r soit aussi un arbre enraciné avec des branches portées), T doit être l'union d'exactly trois chemins portés $P(r, x_1), P(r, x_2)$ et $P(r, x_3)$. En d'autres mots, x_1, x_2 et x_3 doivent constituer l'ensemble des feuilles de T .

Soit y_i le voisin de x_i dans le chemin $P(r, x_i)$ ($i = 1, 2, 3$). Comme G est biparti, soit $x_i \in I(y_i, u)$, soit $y_i \in I(x_i, u)$. Mais comme $x_i \in \Upsilon(u, T)$, nous avons $x_i \in I(y_i, u)$. Soit T'_i le sous-arbre de T obtenu en supprimant la feuille x_i . De par l'hypothèse de minimalité de T , nous ne pouvons pas remplacer T par son sous-arbre T'_i . Autrement dit, $|\Upsilon(u, T'_i)| \leq 2$. Puisque $x_j, x_k \in \Upsilon(u, T'_i)$ pour $\{i, j, k\} = \{1, 2, 3\}$, x_j ou x_k doit nécessairement appartenir à $I(y_i, u)$ (c.-à-d., $I(y_i, u) \cap \{x_j, x_k\} \neq \emptyset$).

Remarquons que $x_1, x_2, x_3 \in I(u, r)$. En effet, dénotons par z_i le médian du triplet x_i, u, r . Comme $z_i \in I(x_i, r) = P(x_i, r) \subseteq T$ et $z_i \in I(u, x_i)$, si $z_i \neq x_i$, alors nous obtenons une contradiction avec l'inclusion de x_i dans $\Upsilon(u, T)$. Donc $z_i = x_i$, et $x_i \in I(u, z_i)$.

Nous pouvons maintenant supposer, sans perte de généralité, que $d_G(r, x_3) = \max\{d_G(r, x_i) : i = 1, 2, 3\} =: k$. Nous pouvons aussi supposer, toujours sans perte de généralité, que $x_2 \in I(y_3, u)$ car $I(y_3, u) \cap \{x_1, x_2\} \neq \emptyset$ comme nous l'avons expliqué plus haut. Vu que $x_3 \in I(y_3, u)$, nous obtenons que $d_G(x_3, u) + 1 = d_G(y_3, x_2) + d_G(x_2, u)$. Donc $d_G(x_3, u) \geq d_G(x_2, u)$ et nous pouvons conclure que $d_G(x_3, u) = d_G(x_2, u)$ et $d_G(y_3, x_2) = 1$ (c.-à-d., $y_2 = y_3 =: y$). Et comme $x_2, x_3 \in I(r, u)$, nous obtenons aussi que $d_G(x_3, r) = d_G(x_2, r)$. Deux cas se présentent alors :

Cas 1. $d_G(x_1, r) = k$.

Puisque chacun des trois sommets x_1, x_2 et x_3 est à la même distance k de la racine r , nous pouvons appliquer à x_1 le même raisonnement que celui fait pour x_3 afin de déduire que le voisin y_1 de x_1 dans T coïncide avec y_2 ou y_3 . Étant donné que $y_2 = y_3 = y$, nous pouvons conclure que les sommets x_1, x_2 et x_3 ont tous pour voisin y dans T . Vu que y est plus proche de r que ne le sont x_1, x_2 et x_3 , et comme $x_1, x_2, x_3 \in I(r, u)$, nous obtenons que $x_1, x_2, x_3 \in I(y, u)$. En appliquant trois fois la condition du quadrangle, nous trouvons trois sommets $x_{i,j}, i, j \in \{1, 2, 3\}, i \neq j$, tels que $x_{i,j} \sim x_i, x_j$ et $d_G(x_{i,j}, u) = k - 1$ (voir la figure 7.5, gauche). Si deux des sommets $x_{1,2}, x_{2,3}$ et $x_{1,3}$ coïncident, alors nous obtenons un $K_{2,3}$, interdit car c'est un sous-graphe interdit des graphes médians. Les sommets $x_{1,2}, x_{2,3}$ et $x_{1,3}$ sont alors deux-à-deux distincts. Comme G est biparti, cela implique que $d_G(x_i, x_{j,k}) = 3$ pour $\{i, j, k\} = \{1, 2, 3\}$. Par une nouvelle application de la condition du quadrangle à $(x_{1,2}, x_{2,3}, x_2, u)$, nous dérivons un sommet w tel que $w \sim x_{1,2}, x_{2,3}$ et $d_G(w, u) = k - 2$. Comme G est biparti, $d_G(w, x_{1,3})$ vaut soit 3, soit 1. Si $d_G(w, x_{1,3}) = 3 = d(y, w)$, alors le triplet $y, w, x_{1,3}$ possède deux médians x_1 et x_3 , ce qui n'est pas possible car G est un graphe médian. Nous avons, par conséquent, $d_G(w, x_{1,3}) = 1$ (c.-à-d., $w \sim x_{1,3}$). Dans ce cas, les sommets $y, x_1, x_2, x_3, x_{1,2}, x_{2,3}, x_{1,3}, w$ définissent un 3-cube isométrique de G , ce qui est en contradiction avec le fait que G soit sans cube. Cela termine l'analyse du premier cas.

Cas 2. $d_G(x_1, r) < k$.

Ce cas implique que $d_G(r, x_1) \leq k - 1 = d_G(r, y)$. Soit r' le voisin de r sur le (r, y) -chemin de T . Remarquons que $r' \notin I(r, x_1) = P(r, x_1)$. En effet, dans le cas contraire, nous aurions $r' \in P(r, x_1) \cap P(r, x_2) \cap P(r, x_3)$ et nous pourrions remplacer l'arbre T par son sous-arbre T' enraciné en r' consistant en l'union des sous-chemins de $P(r, x_i)$ entre r' et x_i , pour $i = 1, 2, 3$. T' serait alors clairement un arbre enraciné, avec des branches portées, et tel que $x_1, x_2, x_3 \in \Upsilon(u, T')$, contredisant le choix minimal du contre-exemple T . Donc $r' \notin P(r, x_1)$.

Posons $P(r, x_1) = (r, v_1, \dots, v_{m-1}, v_m =: x_1)$. Notons que r peut coïncider avec y_1 , et que x_1 peut coïncider avec v_1 . Comme $v_1, r' \in I(r, u)$, en appliquant la condition du quadrangle, nous trouvons un sommet $v'_2 \sim v_1, r'$ à distance $d_G(r, u) - 2$ de u . Comme $r' \notin I(r, x_1)$, $v'_2 \neq v_2$. Puisque $v_2, v'_2 \in I(v_1, u)$, par la

condition du quadrangle, nous dérivons $v'_3 \sim v_2, v'_2$ à distance $d_G(r, u) - 3$ de u . Une fois de plus, comme $r' \notin I(r, x_1)$, $v'_3 \neq v_3$. En continuant ainsi, nous dérivons des sommets $v'_2, v'_3, \dots, v'_m, v'_{m+1} =: x'_1$ formant un (r', x'_1) -chemin $P(r', x'_1)$ et tels que $v'_{i+1} \sim v_i, v'_i, v'_{i+1} \neq v_{i+1}$ et v'_{i+1} soit d'un cran plus proche de u que v_i et v'_i (voir la figure 7.5, droite). De par sa construction, le chemin $P(r', x'_1)$ est un plus court chemin, dont nous affirmons qu'il est porté. En effet, si ce n'est pas le cas, nous pouvons appliquer le lemme 6 et utiliser le fait que $P(r', x'_1)$ soit un plus court chemin pour trouver deux sommets v'_{i-1}, v'_{i+1} possédant un voisin commun z' différent de v'_i . Soit z le médians du triplet z', v_{i-1}, v_{i+1} . Alors z est un voisin commun de z', v_{i-1} et v_{i+1} , et z est différent de v_i (autrement, nous obtenons un sous-graphe interdit $K_{2,3}$). Mais alors, il est facile de vérifier que les sommets $v_{i-1}, v_i, v_{i+1}, v'_{i-1}, v'_i, v'_{i+1}, z$ et z' induisent un 3-cube isométrique, contrairement à l'hypothèse $\dim(G) \leq 2$. Par conséquent, $P(r', x'_1)$ est bien un chemin porté de G .

Soit T'' l'arbre enraciné en r' consistant en le chemin porté $P(r', x'_1)$ et les sous-chemin portés de $P(r, x_2)$ et $P(r, x_3)$ compris respectivement entre r' et x_2 et entre r' et x_3 . T'' est évidemment un arbre enraciné avec des branches portées. Remarquons que $x'_1, x_2, x_3 \in \Upsilon(u, T'')$. En effet, si x_2 ou x_3 appartenait à $I(x'_1, u)$, alors x'_1 appartiendrait à $I(x_1, u)$ et nous conclurions que x_2 ou x_3 appartient à $I(x_1, u)$, ce qui est impossible car $x_1 \in \Upsilon(u, T)$. D'autre part, x'_1 ne peut appartenir ni à $I(x_2, u)$, ni à $I(x_3, u)$ car $d_G(x'_1, u) = d_G(x_1, u) - 1 \leq d_G(x_2, u) = d_G(x_3, u)$. Par conséquent, $|\Upsilon(u, T'')| \geq 3$. Comme T'' contient moins de sommets que T , nous obtenons une contradiction avec la minimalité de T . Cela conclut l'analyse du second cas et la preuve que T est quasi-porté. \square

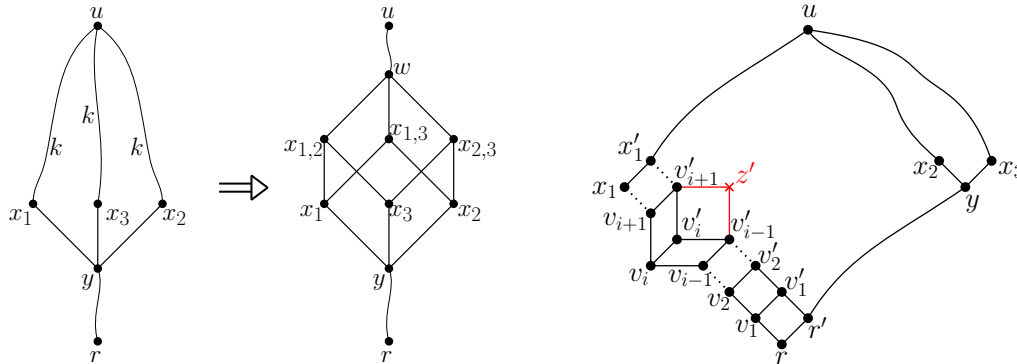


FIGURE 7.5 – Illustration des cas 1 et 2 de la preuve du lemme 32.

En appliquant les lemmes 31 et 32 au sous-graphe médian sans cube de G induit par la fibre $F(x)$, nous obtenons directement :

Corollaire 9. *La bordure $\partial^* F(x)$ de toute fibre $F(x)$ est quasi-portée.*

7.6 Classification des paires de sommets et plus courts chemins

Dans la sous-section 7.5.1 nous avons classifié les fibres de $\text{St}(z)$ en cônes et panneaux. Dans la présente section, nous utilisons cette classification pour distinguer des paires de sommets de G respectivement à sa partition en fibres. Cela étend la classification faite par CHEPOI, DRAGAN et VAXÈS [53] pour les graphes médians planaires.

Soit z un sommet quelconque fixé d'un graphe médian sans cube $G = (V, E)$. Soit $\mathcal{F}_z := \{F(x) : x \in \text{St}(z)\}$ la partition de V en les fibres de $\text{St}(z)$. Soient u et v deux sommets arbitraires de G supposés appartenir aux fibres $F(x)$ et $F(y)$ de \mathcal{F}_z , respectivement. Nous disons que u et v sont *proches* s'ils appartiennent à la même fibre (c.-à-d., $x = y$). Lorsque $F(x)$ et $F(y)$ sont voisines (alors l'une des fibres est un panneau et l'autre un cône), nous disons que u et v sont *1-voisins*. Quand $F(x)$ et $F(y)$ sont deux cônes distincts et voisins d'un même panneau (c.-à-d., $\exists w \in \text{St}(z)$ t.q. $F(w) \sim F(x), F(y)$), u et v seront dits *2-voisins*. Enfin, dans tous les autres cas (c.-à-d., $F(x)$ et $F(y)$ sont distincts, non-voisins et si ce sont deux cônes, alors ils n'ont pas de panneau voisin commun), u et v seront dits *éloignés*. Cette classification des paires de sommets est illustrée dans la figure 7.6. Il n'est pas difficile de constater que ces quatre définitions classifient bien toutes les paires de sommets de G . Nous pouvons aussi remarquer que cette classification possède la propriété suivante : si u' appartient à la même fibre $F(x)$ que u et v' appartient à la même fibre $F(y)$ que v , alors la paire (u', v') est dans la même classe que la paire (u, v) .

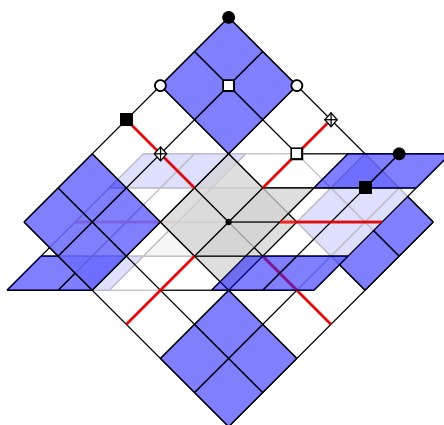


FIGURE 7.6 – Illustration de l'étoile $\text{St}(z)$ (en gris) d'un sommet z , de ses fibres (les cônes sont en bleu et les panneaux en rouge) et de la classification des paires de sommets : les cercles vides sont proches, les cercles pleins sont 2-voisins, les carrés vides sont 1-voisins et les carrés pleins ainsi que les losanges "croisés" sont éloignés.

Nous continuons avec des formules précisant les distances entre les sommets éloignés, 2-voisins et 1-voisins. Chacune de ces formules est illustrée dans la figure 7.7.

Lemme 33. *Pour deux sommets u et v appartenant respectivement aux fibres $F(x)$ et $F(y)$ de $\text{St}(z)$, les conditions suivantes sont équivalentes.*

- (i) u et v sont éloignés ;
- (ii) $I(x, z) \cap I(y, z) = \{z\}$;
- (iii) $d_G(u, v) = d_G(u, z) + d_G(z, v)$ (c.-à-d., $z \in I(u, v)$).

Démonstration. (i) \iff (ii) : Remarquons que u et v sont éloignés si et seulement si $x \neq y$ et ou bien $F(x)$ et $F(y)$ sont deux panneaux, ou bien ce sont deux cônes n'ayant aucun panneau voisin en commun, ou bien l'un est un cône et l'autre un panneau qui ne lui est pas voisin. Si $F(x)$ et $F(y)$ sont deux panneaux, alors $I(x, z) = \{x, z\}$ et $I(y, z) = \{y, z\}$, et donc $I(x, z) \cap I(y, z) = \{z\}$. Si $F(x)$ et $F(y)$ sont deux cônes, alors $I(x, z)$ et $I(y, z)$ sont deux carrés Q_x et Q_y . Par le lemme 27, Q_x et Q_y s'intersectent sur une arête wz si et seulement si $F(w)$ est un panneau voisin de $F(x)$ et de $F(y)$ (c.-à-d., si et seulement si u et v ne sont pas éloignés). Enfin, si $F(x)$ est un cône et $F(y)$ est un panneau, alors $I(x, z)$ est un carré Q_x et $I(y, z)$ est une arête yz . Dans ce cas, $F(x)$ et $F(y)$ ne sont pas voisins si et seulement si l'arête yz n'est pas une arête du carré Q_x (c.-à-d., si et seulement si $I(x, z) \cap I(y, z) = \{z\}$).

(ii) \iff (iii) : Supposons pour commencer que $I(x, z) \cap I(y, z) = \{z\}$. Pour montrer que $z \in I(u, v)$, il suffit de montrer que z est le médian du triplet (u, v, z) . Supposons, par l'absurde, que le médian de u, v et z soit un sommet w distinct de z . Soit s le voisin de z dans $I(z, w)$. Alors clairement, $s \in \text{St}(z)$. Comme $I(x, z) \cap I(y, z) = \{z\}$, s ne peut appartenir aux deux intervalles $I(x, z)$ et $I(y, z)$ simultanément. Disons que $s \notin I(x, z)$. Cela implique que $d_G(s, x) = d_G(z, x) + 1$. Étant donné que x est la porte de u sur $\text{St}(z)$ et que $s \in \text{St}(z)$, nous avons forcément $x \in I(u, s)$. Cela implique qu'il existe un plus court (s, u) -chemin passant par z et x (c.-à-d., $d_G(s, u) = 1 + d_G(z, u)$). D'autre part, puisque $s \in I(z, w) \subset I(z, u)$, nous pouvons conclure que $d_G(z, u) = 1 + d_G(s, u)$. En comparant les deux inégalités, nous obtenons une contradiction. Cela établit que (ii) \implies (iii).

Réciproquement, supposons que $z \in I(u, v)$. Cela implique que z est le médian du triplet (u, v, z) et que $I(u, z) \cap I(v, z) = \{z\}$. Comme x est la porte de u et y est la porte de v sur $\text{St}(z)$, nous concluons que $x \in I(u, z)$ et $y \in I(v, z)$. Par conséquent, $I(x, z) \subseteq I(u, z)$ et $I(y, z) \subseteq I(v, z)$, prouvant que $I(x, z) \cap I(y, z) = \{z\}$. Cela démontre que (iii) \implies (ii). \square

Remarque 23. L'équivalence (ii) \iff (iii) du lemme 33 est aussi vérifiée dans le cas des graphes médians quelconques.

Lemme 34. Soient u et v deux sommets 1-voisins tels que u appartienne à un panneau $F(x)$ et v appartienne à un cône $F(y)$. Soit u_1 et u_2 les deux empreintes de u sur la bordure $\partial^*F(x)$ (il se peut que $u_1 = u_2$) et soit v^+ la porte de v sur $F(x)$. Alors,

$$d_G(u, v) = \min\{d_G(u, u_1) + d_{\partial^*F(x)}(u_1, v^+), d_G(u, u_2) + d_{\partial^*F(x)}(u_2, v^+)\} + d_G(v^+, v).$$

Démonstration. Du fait du lemme 26, $F(x)$ est porté. Par conséquent, il doit exister un plus court (u, v) -chemin passant par v^+ . Les sommets u_1 , u_2 et v^+ appartiennent à la bordure $\partial^*F(x)$ of $F(x)$. Comme, par le lemme 29, $\partial^*F(x)$ est un arbre isométrique et comme, par le lemme 32, u possède au plus deux empreintes u_1 et u_2 sur $\partial^*F(x)$, nous concluons que $d_G(u, v^+) = \min\{d_G(u, u_1) + d_{\partial^*F(x)}(u_1, v^+), d_G(u, u_2) + d_{\partial^*F(x)}(u_2, v^+)\}$. Ainsi, il existe un plus court (u, v) -chemin passant d'abord par l'un des sommets u_1 et u_2 puis par v^+ , montrant l'égalité annoncée. \square

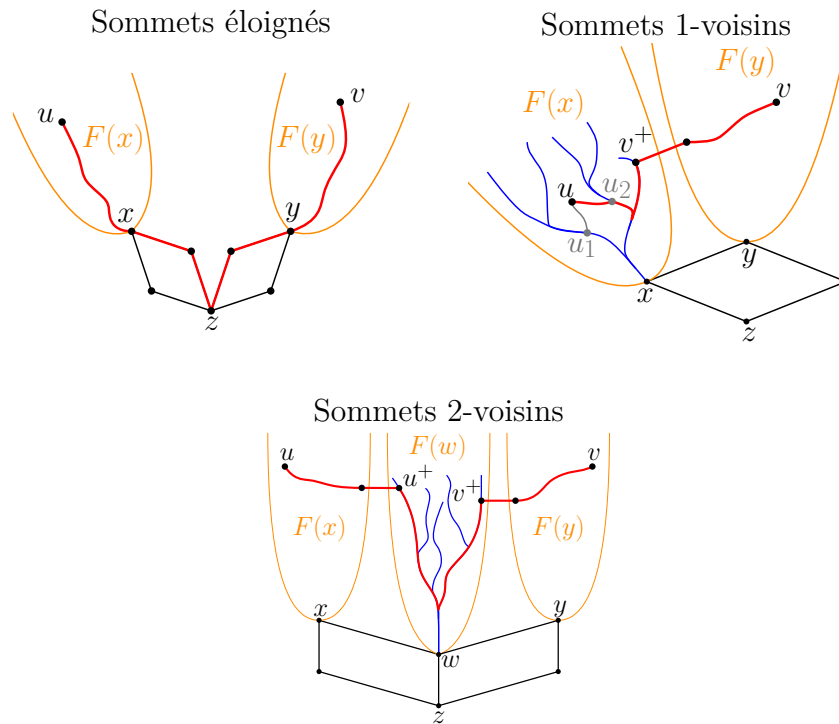


FIGURE 7.7 – Illustration des lemmes 33, 34 et 35 : exemples schématiques de plus courts chemins (en rouge) entre sommets u et v éloignés, 1-voisins et 2-voisins. La bordure des panneaux est dessinée en bleu.

Lemme 35. Soient u et v deux sommets 2-voisins appartenant respectivement aux cônes $F(x)$ et $F(y)$, et soit $F(w)$ le panneau voisin de $F(x)$ et $F(y)$. Soit u^+ et v^+

les portes de u et de v sur $F(w)$, respectivement. Alors

$$d_G(u, v) = d_G(u, u^+) + d_{\partial^* F(w)}(u^+, v^+) + d_G(v^+, v).$$

Démonstration. Étant donné que le demi-espace $W(w, z)$ est convexe et que $u, v \in F(x) \cup F(w) \cup F(y) \subseteq W(w, z)$, tout plus court (u, v) -chemin $P(u, v)$ est contenu dans $W(w, z)$. Nous affirmons que, puisque $u \in F(x)$, $v \in F(y)$ et puisque les fibres $F(x)$ et $F(y)$ ne sont pas voisines, lorsque nous nous déplaçons sur $P(u, v)$ de u vers v , nous devons sortir de $F(x)$ et entrer dans un panneau voisin de $F(x)$. Or, le cône $F(x)$ n'a que deux panneaux voisins : $F(x)$ et un second panneau $F(w') \subseteq W(z, w)$ (c.-à-d., $F(w') \not\subseteq W(w, z)$). Comme $P(u, v) \subseteq W(w, z)$, $P(u, v)$ doit forcément passer par $F(w)$ (et pas par $F(w')$). De même, nous pouvons démontrer que, lorsque nous nous déplaçons de v vers u le long de $P(u, v)$, nous devons aussi quitter $F(y)$ pour passer dans $F(w)$. Ainsi, $P(u, v)$ doit être inclus dans l'union des trois fibres $F(x)$, $F(w)$ et $F(y)$ car ces dernières sont portées.

Nous devons maintenant montrer que u^+ et v^+ appartiennent à un même plus court (u, v) -chemin. En effet, par ce que nous avons montré précédemment, nous savons que tout plus court (u, v) -chemin intersecte $F(w)$. En particulier, il existe un sommet $s \in I(u, v) \cap F(w)$. Comme u^+ est la porte de u sur $F(w)$ et v^+ est la porte de v sur $F(w)$, nous déduisons que $u^+ \in I(u, s)$ et $v^+ \in I(v, s)$. Et, puisque $s \in I(u, v)$, il existe un plus court chemin de u vers v passant par u^+ , s et v^+ . Cela démontre que $d_G(u, v) = d_G(u, u^+) + d_G(u^+, v^+) + d_G(v^+, v)$. Comme $\partial^* F(w)$ est un arbre isométrique, $d_G(u^+, v^+) = d_{\partial^* F(w)}(u^+, v^+)$, établissant l'égalité $d_G(u, v) = d_G(u, u^+) + d_{\partial^* F(w)}(u^+, v^+) + d_G(v^+, v)$. \square

7.7 Schéma de distance pour les graphes médians sans cube

Soit $G = (V, E)$ un graphe médian sans cube avec n sommets et soit m un sommet médian de G . Soient u et v n'importe quels deux sommets de G entre lesquels nous voulons calculer la distance $d_G(u, v)$. En appliquant les lemmes 33, 34 et 35 de la section 7.6 avec m jouant le rôle de z , la distance $d_G(u, v)$ entre u et v peut être calculée dès lors que u et v sont éloignés, 1-voisins ou 2-voisins, à supposer que u et v aient gardé dans leurs étiquettes les distances vers m et leurs portes (u^+ et v^+) ou empreintes (u_1, u_2 et v_1, v_2) respectives, selon que u et v soient dans des cônes ou des panneaux (respectivement). Le calcul de $d_G(u, v)$ nécessite aussi que u et v aient gardé toutes les informations utiles au calcul des distances $d_{\partial^* F(x)}(u_1, v^+)$, $d_{\partial^* F(x)}(u_2, v^+)$ et $d_{\partial^* F(w)}(u^+, v^+)$. Comme les bordures sont des arbres isométriques, cela peut être fait en gardant dans l'étiquette de u des étiquettes de u_1 et u_2 ou de u^+ attribuées par un schéma de distance pour

les arbres (et de même pour les portes ou empreintes de v). Cela montre que $d_G(u, v)$ peut être calculée dans tous les cas, excepté lorsque u et v sont proches. Mais, puisque $F(x)$ est portée, et donc médian, nous pouvons alors appliquer la même procédure récursivement sur chaque fibre $F(x)$. En procédant ainsi, $d_G(u, v)$ peut être calculée la première fois que u et v n'appartiennent plus à une même fibre dans les appels récursifs. Puisqu'à chaque étape, la partition est effectuée par rapport à un sommet médian, le nombre de sommets est au moins réduit de moitié par le lemme 30. Par conséquent, l'arbre des appels récursifs sera de profondeur logarithmique.

Dans cette section, nous donnons la description formelle de notre schéma d'étiquetage de distance. La fonction d'encodage est décrite par l'algorithme [ENC_DIST](#), présenté dans la section 7.7.1. Dans la section 7.7.2, nous présentons l'algorithme [DISTANCE](#) utilisé pour répondre aux requêtes de distances.

7.7.1 Encodage

Soit $G = (V, E)$ un graphe médian sans cube avec n sommets. Nous expliquons ici comment [ENC_DIST](#) construit une étiquette de distance $LD(u)$ pour chaque sommet u de G . Ces étiquettes sont construites via une procédure récursive. Chaque profondeur de la récursion sera appelée une *étape*. Au départ, nous supposons que chaque sommet u de G est distingué par un identifiant unique $id(u)$. Nous nous référerons à cette étape de nommage comme l'*étape 0*, et nous dénoterons la partie correspondante de $LD(u)$ par $LD_0(u)$ (c.-à-d., $LD_0(u) := id(u)$). À l'*étape 1*, [ENC_DIST](#) calcule un sommet médian m de G , l'étoile $St(m)$ de m , ainsi que la partition en fibres $\mathcal{F}_m := \{F(x) : x \in St(m)\}$ de V . Chaque sommet u de G (m inclus) reçoit alors l'identifiant $id(m)$ de m et sa distance $d_G(u, m)$ à m . Après cela, chaque sommet x de $St(m)$ reçoit un identifiant "spécial" $L_{St(m)}(x)$ de longueur $O(\log |V|)$ consistant en un étiquetage de l'étoile $St(m)$, comme décrit dans la section 7.3.1. Ensuite, [ENC_DIST](#) calcule la porte u^\downarrow de chaque sommet u de G sur $St(m)$ et ajoute son identifiant $L_{St(m)}(u^\downarrow)$ à $LD(u)$. Remarquons que les identifiants $L_{St(m)}(x)$ des sommets de $St(m)$ peuvent être utilisés pour distinguer les fibres de $St(m)$. Ce triplet $(id(m), d_G(u, m), L_{St(m)}(u^\downarrow))$ contient toutes les informations nécessaires relatives à $St(m)$. Nous nous référerons donc à cette partie d'étiquette comme la *partie "étoile"* de l'information $LD_1(u)$ donnée au sommet u à l'étape 1. Cette partie sera dénotée par $LD_1^{St}(u)$. Posons aussi $LD_1^{St[Med]}(u) := id(m)$, $LD_1^{St[Dist]}(u) := d_G(u, m)$ et $LD_1^{St[Porte]}(u) := L_{St(m)}(u^\downarrow)$ les trois composantes de la partie d'étiquette $LD_1^{St}(u)$.

Suite à cela, toujours à l'étape 1, l'algorithme considère chaque fibre $F(x)$ de \mathcal{F}_m . Si $F(x)$ est un panneau, alors l'algorithme calcule la bordure $\partial^* F(x)$ de $F(x)$

(arbre isométrique quasi-porté). Les sommets v de cet arbre $\partial^* F(x)$ reçoivent un identifiant spécial $LD_{\partial^* F(x)}(v)$ de longueur $O(\log^2 |V|)$ consistant en une étiquette de distance pour les arbres (par exemple, celles décrites par FREEDMAN et al. [75] autorisant un décodage en temps constant). Pour chaque sommet u d'un panneau $F(x)$, l'algorithme calcule les deux empreintes u_1 et u_2 de u sur $\partial^* F(x)$ (il se peut que $u_1 = u_2$) et stocke $(LD_{\partial^* F(x)}(u_1), d_G(u, u_1))$ et $(LD_{\partial^* F(x)}(u_2), d_G(u, u_2))$ dans $LD_1^{\text{er}}(u)$ et $LD_1^{\text{2nd}}(u)$.

Si $F(x)$ est un cône et $F(w_1)$ et $F(w_2)$ sont deux panneaux voisins de $F(x)$, alors l'algorithme calcule, pour chaque sommet u de $F(x)$, les portes u_1^+ et u_2^+ de u sur $F(w_1)$ et $F(w_2)$, respectivement. Comme $u_i^+ \in \partial_x F(w_i) \subseteq \partial^* F(x)$, $i = 1, 2$, les étiquettes $LD_{\partial^* F(w_1)}(u_1^+)$ et $LD_{\partial^* F(w_2)}(u_2^+)$ données par des schémas de distances sur les arbres $\partial^* F(w_1)$ et $\partial^* F(w_2)$ sont bien définies. Ainsi, l'algorithme stocke $(LD_{\partial^* F(w_1)}(u_1^+), d_G(u, u_1^+))$ et $(LD_{\partial^* F(w_2)}(u_2^+), d_G(u, u_2^+))$ dans $LD_1^{\text{er}}(u)$ et $LD_1^{\text{2nd}}(u)$. Cela conclut l'étape 1.

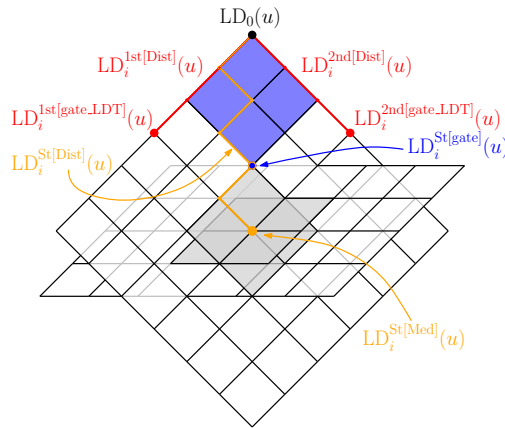


FIGURE 7.8 – Illustration de $LD_0(u)$ et de l'information ajoutée à $LD(u)$ à l'étape i .

Puisque \mathcal{F}_m partitionne V en sous-graphes médians, l'étiquette $LD_2(u)$ ajoutée à $LD(u)$ à l'étape 2 est construite comme $LD_1(u)$, en remplaçant G par la fibre $F(u^\perp)$ contenant u , et ainsi de suite. Comme chaque fibre contient au plus la moitié des sommets du graphe courant, à l'étape $\lceil \log |V| \rceil$, chaque fibre consiste en un unique sommet et l'algorithme s'arrête. Par conséquent, pour chaque paire de sommets u et v de G , il existe une étape de la récursion après laquelle u et v ne sont plus proches. La figure 7.8 donne une illustration des parties de $LD_i(u)$.

Continuons avec le pseudo-code de l'algorithme d'encodage. Pour un vecteur $L(v) := (t_1, \dots, t_k)$ de vecteurs t_1, \dots, t_k et un vecteur arbitraire t , nous dénotons par $L(v) \circ t := (t_1, \dots, t_k, t)$ la *concatenation* de $L(v)$ et t .

Remarque 24. Pour que les algorithmes d'encodage (que ce soit de distance ou de routage plus tard) soient plus lisibles, nous n'y faisons pas la précision suivante.

Si u appartient à un cône et si v et w sont les portes de ses panneaux voisins (à une certaine étape i), alors $\text{LD}_i^{\text{St[Porte]}}(u) = \{\text{LD}_i^{\text{St[Porte]}}(v), \text{LD}_i^{\text{St[Porte]}}(w)\}$. Pour la suite, il sera bien pratique de supposer que les informations dans $\text{LD}_i(u)$ relatives au panneau $F(x)$ ($x \in \{v, w\}$) tel que $\text{LD}_i^{\text{St[Porte]}}(x) = \{\min\{j : j \in \text{LD}_i^{\text{St[Porte]}}(u)\}\}$ sont stockées dans $\text{LD}_i^{\text{1er}}(u)$. Cela économise un champ dans $\text{LD}_i(u)$.

Algorithme 1 : ENC_DIST ($G, \text{LD}(V)$)

Entrée : Un graphe médian sans cube $G = (V, E)$ et un étiquetage $\text{LD}(V)$ de ses sommets consistant, initialement, en un identifiant unique $\text{id}(u)$ donné à chaque sommet $u \in V$

- 1 si $V = \{v\}$ alors stop ;
- 2 Trouver un sommet médian m de G ;
- 3 $L_{\text{St}(m)}(\text{St}(m)) \leftarrow \text{Enc_Etoile_Med}(\text{St}(m))$;
- 4 **pour chaque** panneau $F(x) \in \mathcal{F}_m$ **faire**
- 5 $\text{LD}_{\partial^*F(x)}(\partial^*F(x)) \leftarrow \text{Enc_Dist_Arbre}(\partial^*F(x))$;
- 6 **pour chaque** $u \in F(x)$ **faire**
- 7 Trouver les empreintes u_1 et u_2 de u sur $\partial^*F(x)$;
- 8 $L_{\text{St}} \leftarrow (\text{id}(m), d_G(u, m), L_{\text{St}(m)}(x))$;
- 9 $L_{\text{1er}} \leftarrow (\text{LD}_{\partial^*F(x)}(u_1), d_G(u, u_1))$;
- 10 $L_{\text{2nd}} \leftarrow (\text{LD}_{\partial^*F(x)}(u_2), d_G(u, u_2))$;
- 11 $\text{LD}(u) \leftarrow \text{LD}(u) \circ (L_{\text{St}}, L_{\text{1er}}, L_{\text{2nd}})$;
- 12 ENC_DIST ($F(x), \text{LD}(V)$) ;
- 13 **pour chaque** cône $F(x) \in \mathcal{F}_m$ **faire**
- 14 **pour chaque** $u \in F(x)$ **faire**
- 15 Trouver les panneaux $F(w_1)$ et $F(w_2)$ voisins de $F(x)$;
- 16 Trouver les portes u_1^+ et u_2^+ de u sur $F(w_1)$ et $F(w_2)$;
- 17 $L_{\text{St}} \leftarrow (\text{id}(m), d_G(u, m), L_{\text{St}(m)}(x))$;
- 18 $L_{\text{1er}} \leftarrow (\text{LD}_{\partial^*F(w_1)}(u_1^+), d_G(u, u_1^+))$;
- 19 $L_{\text{2nd}} \leftarrow (\text{LD}_{\partial^*F(w_2)}(u_2^+), d_G(u, u_2^+))$;
- 20 $\text{LD}(u) \leftarrow \text{LD}(u) \circ (L_{\text{St}}, L_{\text{1er}}, L_{\text{2nd}})$;
- 21 ENC_DIST ($F(x), \text{LD}(V)$) ;

7.7.2 Requêtes de distance

Soient u et v deux sommets d'un graphe médian sans cube $G = (V, E)$ et soient $\text{LD}(u)$ et $\text{LD}(v)$ leurs étiquettes, données par **ENC_DIST**. Nous décrivons ici comment l'algorithme de décodage **DISTANCE** calcule les positions relatives de u et v l'un par rapport à l'autre (c.-à-d., le statut proches, éloignés, 1-voisins ou 2-voisins de (u, v)) et utilise cela pour calculer la distance exacte $d_G(u, v)$.

7.7.2.1 L'algorithme

Nous continuons avec la présentation formelle de l'algorithme **DISTANCE**. Les fonctions `Dist_1-Voisins`, `Dist_2-Voisins` et `Dist_Eloignes`, utilisées dans cet algorithme, sont décrites dans la section 7.7.2.2 suivante. La fonction `Dist_Etoile` est celle décrite en section 7.3.1.

Algorithme 2 : `DISTANCE` ($LD(u)$, $LD(v)$)

Entrée : Les étiquettes $LD(u)$ et $LD(v)$ de deux sommets u et v de G
 Sortie : La distance entre u et v dans G

- 1 si $LD_0(u) = LD_0(v)$ /* $u = v$ */ alors retourner 0 ;
- 2 Soit i le plus grand entier tel que $LD_i^{St[Med]}(u) = LD_i^{St[Med]}(v)$;
- 3 $d \leftarrow \text{Dist_Etoile}(LD_i^{St[Porte]}(u), LD_i^{St[Porte]}(v))$; // $d_G(u^\downarrow, v^\downarrow)$
- 4 $d_u \leftarrow \text{Dist_Etoile}(LD_i^{St[Porte]}(u), 0)$; // $d_G(u^\downarrow, m)$
- 5 $d_v \leftarrow \text{Dist_Etoile}(LD_i^{St[Porte]}(v), 0)$; // $d_G(v^\downarrow, m)$
- 6 si $d = 1$ et $d_u = 1$ alors retourner `Dist_1-Voisins`($LD_i(u)$, $LD_i(v)$);
- 7 si $d = 1$ et $d_v = 1$ alors retourner `Dist_1-Voisins`($LD_i(v)$, $LD_i(u)$);
- 8 si $d = 2$ et $d_u = d_v = 2$ alors retourner `Dist_2-Voisins`($LD_i(u)$, $LD_i(v)$);
- 9 retourner `Dist_Eloignes`($LD_i(u)$, $LD_i(v)$) ;

7.7.2.2 Descriptions et fonctions

Étant donnés les étiquettes $LD(u)$ et $LD(v)$ de deux sommets u et v , l'algorithme de décodage commence par tester si u et v coïncident. C'est ce qui est fait en ligne 1 de **DISTANCE**. Si $u \neq v$, alors **DISTANCE** cherche le plus grand entier i tel que $LD_i^{St[Med]}(u) = LD_i^{St[Med]}(v)$ (ligne 2). Cela correspond à la première fois que les sommets u et v ont appartenu à des fibres différentes dans la partition d'un même graphe. Soit m le sommet médian du graphe médian courant. Par souci de concision, nous utiliserons parfois l'abus de langage "médians séparateur" de u et v pour parler de ce sommet m (bien qu'il ne sépare pas u et v mais les rende 1-voisins, 2-voisins ou éloignés). Dans les lignes 3, 4 et 5, l'algorithme **DISTANCE** calcule les distances d , d_u et d_v entre les portes u^\downarrow et v^\downarrow de u et v dans l'étoile $St(m)$, et les distances de u^\downarrow et v^\downarrow à m , respectivement. Ce calcul est effectué en utilisant les identifiants $LD_i^{St[Porte]}(u)$ et $LD_i^{St[Porte]}(v)$ et le décodeur de distance pour les étoiles. Nous pouvons facilement remarquer qu'avec ces informations, il est possible de déterminer, pour chacun des sommets u et v , s'il appartient à un cône ou un panneau et, par la même occasion, de déterminer si u et v sont 1-voisins, 2-voisins ou éloignés. Dans chacun de ces cas, un appel à la fonction appropriée est effectué entre les lignes 6 et 9.

Commençons en supposant que les sommets u et v sont 1-voisins ($d = 1$ et l'un

des d_u, d_v vaut 1 et l'autre 2). Autrement dit, l'un des sommets appartient à un cône et l'autre à un panneau voisin de ce cône. La fonction `Dist_1-Voisins` retourne la distance $d_G(u, v)$ en supposant que u est dans un panneau et v est dans un cône (si v appartient à un panneau et u à un cône, il suffit d'échanger u et v dans l'appel de la fonction `Dist_1-Voisins`). La fonction trouve la porte v^+ de v sur le panneau contenant u en regardant $LD_i^{\text{St[Porte]}}(v)$ (elle récupère au passage la distance $d_G(v, v^+)$). Ensuite, elle récupère l'empreinte u^* de u (et la distance $d_G(u, u^*)$) sur la bordure du panneau qui minimise la distance entre u et l'une de ses deux empreintes additionnée à la distance entre cette empreinte et la porte v^+ en utilisant les étiquettes de distance dans l'arbre bordure. Enfin, `Dist_1-Voisins` retourne $d_G(u, u^*) + d_G(u^*, v^+) + d_G(v^+, v)$ en tant que $d_G(v, u)$.

```

fonction Dist_1-Voisins( $LD_i(u), LD_i(v)$ ):
  dir ← 1er ; // remarque 24
  si  $LD_i^{\text{St[Porte]}}(u) = \min\{j : j \in LD_i^{\text{St[Porte]}}(v)\}$  alors
    dir ← 2nd ;
  // distance entre la porte de  $v$  et la 1ère empreinte de  $u$ 
   $d_1 \leftarrow \text{Dist\_Arbre}(LD_i^{\text{dir[Porte\_LDT]}}(v), LD_i^{\text{1er[Emp\_LDT]}}(u))$  ;
  // distance entre la porte de  $v$  et la 2ème empreinte de  $u$ 
   $d_2 \leftarrow \text{Dist\_Arbre}(LD_i^{\text{dir[Porte\_LDT]}}(v), LD_i^{\text{2nd[Emp\_LDT]}}(u))$  ;
  retourner  $LD_i^{\text{dir[Dist]}}(v) + \min\{d_1 + LD_i^{\text{1er[Dist]}}(u), d_2 + LD_i^{\text{2nd[Dist]}}(u)\}$  ;

```

Supposons maintenant que les sommets u et v soient 2-voisins (c.-à-d., $d = d_u = d_v = 2$). Alors u et v appartiennent tous deux à un cône. En inspectant $LD_i^{\text{St[Porte]}}(u)$ et $LD_i^{\text{St[Porte]}}(v)$, la fonction `Dist_2-Voisins` détermine le panneau $F(w)$ voisin commun des cônes respectifs $F(u^\downarrow)$ et $F(v^\downarrow)$ de u et de v . Ensuite, la fonction récupère les portes respectives u^+ de v^+ de u et v sur le panneau $F(w)$ ainsi que les distances $d_G(u, u^+)$ et $d_G(v, v^+)$. La distance entre les portes u^+ et v^+ est retrouvée en utilisant le décodeur de distance dans les arbres. Au final, l'algorithme répond $d_G(u, u^+) + d_G(u^+, v^+) + d_G(v^+, v)$ à la requête de $d_G(u, v)$.

```

fonction Dist_2-Voisins( $LD_i(u), LD_i(v)$ ):
  pour chaque  $x \in \{u, v\}$  faire
    // le panneau commun est le 1er du cône de  $x$ 
    dir $x$  ← 1er ;
    si  $LD_i^{\text{St[Porte]}}(u) \cap LD_i^{\text{St[Porte]}}(v) = \min\{i : i \in LD_i^{\text{St[Porte]}}(x)\}$  alors
      // le panneau commun est le 2nd du cône de  $x$ 
      dir $x$  ← 2nd ;
   $d \leftarrow \text{Dist\_Arbre}(LD_i^{\text{dir}_u[\text{Porte\_LDT}]}}(u), LD_i^{\text{dir}_v[\text{Porte\_LDT}]}}(v))$  ;
  retourner  $LD_i^{\text{dir}_u[\text{Dist]}}(u) + LD_i^{\text{dir}_v[\text{Dist]}}(v) + d$  ;

```

Dans les cas restants, les sommets u et v sont éloignés. Par le lemme 33, il existe un plus court chemin entre u et v passant par m . Comme u et v ont tous les deux stocké l'identifiant de m ainsi que leur distance à celui-ci, `Dist_Eloignes`

retourne simplement la somme de ces deux distances.

```

fonction Dist_Eloignes( $LD_i(u)$ ,  $LD_i(v)$ ):
    retourner  $LD_i^{\text{St}[\text{Dist}]}(u) + LD_i^{\text{St}[\text{Dist}]}(v)$ ;
    
```

7.7.3 Correction et complexité

La correction de l'algorithme `ENC_DIST` résulte des propriétés suivantes des graphes médians sans cube : les étoiles et les fibres sont portées (lemmes 25 et 26) ; les bordures des fibres sont quasi-portées (corollaire 9) et ce sont des arbres isométriques aux branches portées (lemme 31) ; et des formules pour calculées les distances entre sommets éloignés, 1-voisins et 2-voisins (lemmes 33, 34 et 35).

Nous étudions maintenant la longueur des étiquettes attribuées par `ENC_DIST` et la complexité en temps de leur construction par l'encodage et de leur décodage. Pour étudier la complexité en temps de l'encodage d'un graphe G , nous supposons que nous avons accès à la matrice des distances de G (sinon, la construire prend un temps $O(n^3)$ avec l'algorithme de Floyd-Warshall, par exemple).

Lemme 36. `ENC_DIST` s'exécute sur un graphe médian sans cube $G = (V, E)$ en temps $O(|V|^2 \log |V|)$.

Démonstration. Calculer un sommet médian d'un graphe G avec n sommets, en supposant accessible sa matrice de distances, peut être fait en temps $O(n^2)$. L'encodage d'un sous-arbre de G avec $O(n)$ sommets en utilisant la procédure `Enc_Dist_Arbre` requiert $O(n \log n)$ opérations. Une étoile de G peut avoir une taille de $O(\min\{\Delta(G)^2, n\})$, l'encoder avec `Enc_Etoile_Med` nécessite un temps linéaire en n . Trouver les portes/empreintes de tout sommet de G sur tout sous-graphe porté/quasi-porté H de G de taille $O(n)$ se fait en temps $O(n^2)$. Tout compte fait, un unique appel de `ENC_DIST` prend un temps quadratique en la taille de G . Dénotons par p_i le nombre de parties créées jusqu'à l'étape i , et dénotons par $n_{i,j}$ le nombre de sommets apparaissant dans la j -ième partie créée à l'étape i . Pour chaque étape i ,

$$\sum_{k=1}^{p_i} n_{i,k} = n - p_i = O(n) \quad (7.1)$$

Par conséquent, la complexité en temps d'une étape vaut $O(\sum_{k=1}^{p_i} n_{i,k}^2)$. Comme $\sum_{k=1}^{p_i} n_{i,k}^2 \leq (\sum_{k=1}^{p_i} n_{i,k})^2 = O(n^2)$, et comme le nombre d'étapes est $O(\log n)$ par le lemme 30, `ENC_DIST` a une complexité en temps totale de $O(n^2 \log n)$. \square

Lemme 37. `ENC_DIST` donne à chaque sommet d'un graphe médian sans cube $G = (V, E)$ des étiquettes de longueur $O(\log^3 |V|)$.

Démonstration. À chaque étape de partitionnement, nous sélectionnons un sommet médian. Cela assure, d’après le lemme 30, que chaque sommet $v \in V$ apparaît dans au plus $\lceil \log |V| \rceil$ fibres différentes. Dans chacune de ces fibres, $\text{LD}(v)$ se voit ajouter $O(\log^2 n)$ nouveaux bits. En effet, les informations stockées correspondent aux lignes 8, 9 et 10 (ou 17, 18 et 19) de l’algorithme 1. L_{St} a clairement longueur $O(\log n)$ car c’est le cas de $L_{\text{St}(m)}(v^\downarrow)$, comme vu dans la section 7.3.1. $L_{1\text{er}}$ et $L_{2\text{nd}}$ ont tous deux longueur $O(\log^2 n)$ car les étiquettes de distance pour les arbres qu’ils contiennent ont longueur $O(\log^2 n)$ (FREEDMAN et al. [75]). \square

Pour le décodage des étiquettes, il suffit de lire ces dernières jusqu’à trouver le médian à partir duquel u et v ne sont plus proches. Cela peut être fait en temps $O(\log^2 n)$ (en considérant un modèle RAM). Ensuite, il peut être nécessaire de décoder les étiquettes d’un schéma de distance pour les arbres, mais nécessite un temps constant avec le schéma de FREEDMAN et al. [75].

En résumé, la lecture de $\text{LD}(u)$ et $\text{LD}(v)$ jusqu’à trouver le sommet médian m séparateur de u et v semble constituer l’étape la plus coûteuse du décodage. En fait, avec une structure de donnée appropriée utilisant $O(\log^2 n)$ bits supplémentaires d’information à concaténer à $\text{LD}(u)$ et $\text{LD}(v)$, il est possible de trouver m en temps constant, puis de sauter directement aux parties correspondantes des étiquettes. Pour cela, considérons l’arbre T (des *séparateurs récursifs*) dans lequel les sommets à profondeur i correspondent aux médians choisis dans les parties de l’étape i , dans lequel les fils d’un sommet x sont les médians choisis à l’étape $i + 1$ dans la fibre générée par x à l’étape i . Nous pouvons remarquer que chaque sommet de G apparaît exactement une fois dans cet arbre, que le séparateur m de deux sommets u et v est leur plus proche ancêtre commun dans T , et que la profondeur j de ce sommet correspond à sa position dans $\text{LD}(u)$ et $\text{LD}(v)$ (c.-à-d., $\text{LD}_j^{\text{St}[\text{Med}]}(u) = \text{LD}_j^{\text{St}[\text{Med}]}(v) = \text{id}(m)$). Comme cela a été pointé par PELEG [132], n’importe quel schéma de distance pour les arbres peut être étendu à un schéma de *profondeur du plus proche ancêtre commun* (utilisons l’acronyme anglais *NCAD*, pour “*nearest common ancestor’s depth*”) en ajoutant à l’étiquette de distance $L(u)$ de chaque sommet $u \in V(T)$ sa profondeur $\text{prof}(u)$ dans T . Étant donnés deux sommets u et v de T , le décodeur de *NCAD* peut alors retourner $\frac{1}{2}(\text{prof}(u) + \text{prof}(v) - d_T(u, v))$. Ce calcul requiert alors un temps du même ordre de grandeur que le décodage de $L(u)$ et $L(v)$, ce qui est fait en temps constant (FREEDMAN et al. [75]). Ainsi, pendant l’exécution de `ENC_DIST`, nous pouvons aussi construire l’arbre T des appels récursifs et donner à chaque sommet $u \in V(G)$ une étiquette de *NCAD* $L'(u)$ dans T (puis la concaténer à $\text{LD}(u)$). Maintenant, la première étape de `DISTANCE` consistera à décoder $L'(u)$ et $L'(v)$ (en temps constant). Ensuite, il suffira de lire directement les parties de $\text{LD}(u)$ et $\text{LD}(v)$ correspondant au retour de ce décodage préalable.

Le fait que `DISTANCE` ($\text{LD}(u)$, $\text{LD}(v)$) retourne la distance exacte $d_G(u, v)$ entre u et v découle des lemmes 33, 34 et 35. Le théorème 20 est donc vérifié.

7.8 Schéma de routage pour les graphes médians sans cube

Dans cette section, nous adaptons le schéma d'étiquetage de distance pour obtenir un schéma d'étiquetage de routage pour les graphes médians sans cube. Rappelons qu'un schéma de routage attribue des étiquettes aux sommets d'un graphe de sorte que, étant données les étiquettes d'une source u et d'une destination v , le décodage puisse calculer $\text{port}(u, v)$ (c.-à-d., la direction vers un sommet $u' \in N(u) \cap I(u, v)$).

7.8.1 L'idée

L'idée pour l'encodage dans le cas du routage est la même que celle pour les requêtes de distances de la section 7.7 : le graphe est partitionné récursivement en fibres, respectivement à un sommet médian. À chaque étape, les étiquettes des sommets reçoivent trois nouvelles parties, nommées "St", "1er" et "2nd" comme précédemment. Évidemment, les informations stockées dans ces parties seront un peu différentes de celles utilisées pour décoder des distances. En plus du fait que des informations spécifiques au routage soient nécessaires, cela est aussi dû au fait que les requêtes de routage (contrairement aux requêtes de distance) ne sont pas commutatives. Par exemple, dans la fonction `Dist_1-Voisins`, nous supposons que u appartenait à un panneau et v à un cône. Le cas symétrique était réduit au premier avec un appel à `Dist_1-Voisins` avec v pour premier argument et u pour second. Cela n'est plus possible pour faire du routage car router d'un panneau vers un cône est différent de router d'un cône vers un panneau. L'idée intuitive derrière cela est le fait que, les cônes ayant un nombre constant (égal à deux) de frontières (portées), les sommets à l'intérieur peuvent stocker des informations relatives à chacune des frontières et ainsi connaître les directions vers lesquelles elles mènent. Les panneaux, quant à eux, ont aussi des frontières portées, mais en nombre potentiellement linéaire en la taille du graphe. Pour que les sommets dans ces panneaux stockent un nombre de portes constant, nous devons considérer non pas leurs frontières mais leurs bordures, perdant au passage l'information de direction. Ainsi, lorsqu'une requête de routage d'un panneau vers un cône est effectuée, il faut que les informations relatives aux choix de la direction soient disponibles dans l'étiquette du sommet dans le cône.

Comme dans le schéma de distance, les décisions de routage de u vers v nécessitent au préalable de trouver la premier médian stocké par u et v par rapport auquel les sommets appartiennent à des fibres distinctes. Soit m le sommet médian en question, et soient $F(x)$ et $F(y)$ les fibres de la partition courante par rapport à m contenant u et v , respectivement. Si u et v sont éloignés, alors

$d_G(u, v) = d_G(u, m) + d_G(m, v)$, et donc le routage de u vers v est équivalent au routage de u vers m (à moins que $u = m$). De ce fait, la fonction d'encodage doit donner à l'étiquette de u le port $\text{port}(u, u')$ pour un certain voisin u' de u dans $I(u, m)$. Si $u = m$, alors il suffit de router de $u = m$ vers la porte y de v sur $\text{St}(m)$. Cela est fait en utilisant un schéma de routage pour les étoiles sur $\text{St}(m)$.

Si u et v sont 2-voisins, alors $F(x)$ et $F(y)$ sont deux cônes voisins d'un même panneau $F(w)$. De façon similaire au schéma de distance, le décodage du schéma de routage trouve ce $F(w)$. Puisque les portes u^+ et v^+ de v sur $F(w)$ appartiennent à un même plus court (u, v) -chemin, il suffit de router de u vers u^+ . Ainsi, l'encodage doit ajouter à l'étiquette de u le port de u pour atteindre un voisin dans $I(u, u^+)$. La même information est nécessaire lorsque u et v sont 1-voisins et $F(x)$ est un cône et $F(y)$ un panneau. En effet, dans ce cas, il existe un plus court (u, v) -chemin passant par la porte u^+ de u sur $F(y)$ et par l'une des empreintes de v sur $\partial^*F(y)$. En conséquence, pour router de u vers v , il suffit de router de u vers u^+ .

Enfin, supposons que u et v sont 1-voisins, mais que $F(x)$ soit un panneau et $F(y)$ un cône. Alors, dans ce cas, il existe un plus court (u, v) -chemin passant par l'une des empreintes u_1 ou u_2 de u sur $\partial^*F(x)$ et par la porte v^+ de v sur $F(x)$. Par conséquent, si u est différent de v^+ , il suffit de router le message de u vers l'un de ses voisins dans $I(u, u_1)$ ou dans $I(u, u_2)$ selon la position de v . L'encodage ajoute donc dans l'étiquette de u les ports vers ces deux voisins de u . Pour décider lequel permet de se rapprocher de v , le décodage doit comparer $d_G(u, u_1) + d_G(u_1, v^+)$ et $d_G(u, u_2) + d_G(u_2, v^+)$. Ainsi, à la différence du schéma de routage dans les arbres, le présent schéma de routage dans les graphes médians sans cube doit contenir des informations de distance. D'autre part, si u coïncide avec v^+ , alors le message doit être routé vers un voisin de u dans $I(u, v)$ qui doit forcément appartenir au cône $F(y)$ et non au panneau $F(x)$ (car v^+ est la porte de v sur $F(x)$). Il existe un unique sommet $\text{jum}(v^+)$ de $F(y)$ adjacent à v^+ . Il n'est pas possible de conserver l'identifiant de $\text{jum}(v^+)$ dans l'étiquette de $u = v^+$ car un sommet dans un panneau peut avoir un nombre arbitraire de sommets voisins dans des cônes voisins. L'identifiant de $\text{jum}(v^+)$ doit donc être stocké dans l'étiquette de v (car un cône ne peut avoir que deux panneaux voisins).

7.8.2 Encodage

Nous présentons maintenant l'algorithme d'encodage de façon plus détaillée. Soit $G = (V, E)$ un graphe médian sans cube et soit u un sommet quelconque de G . Soit i une étape récursive dans l'application à G de l'algorithme `ENC_ROUT`, et soit m un sommet médian du sous-graphe médian courant (à l'étape i) contenant u .

La partie “St” $\text{LR}_i^{\text{St}}(u)$ de l’étiquette de u est composée de l’identifiant de m , d’un port à prendre depuis u pour se rapprocher de m , d’un port de m vers u , ainsi que d’une étiquette d’étoile (voir section 7.3.1) de la porte u^\downarrow de u sur $\text{St}(m)$. Remarquons que m ne peut pas stocker un port à prendre pour atteindre chaque autre sommet du graphe. Ainsi, pour pouvoir répondre à des requêtes de routage depuis m , chaque sommet u stocke donc son “port retour” depuis m dans sa partie d’étiquette $\text{LR}_i^{\text{St}[\text{depuisMed}]}(u)$. Suivent les composantes de la partie d’étiquette $\text{LR}_i^{\text{St}}(u)$:

- (1) $\text{LR}_i^{\text{St}[\text{Med}]}(u) := \text{id}(m)$ est l’identifiant unique de m ;
- (2) $\text{LR}_i^{\text{St}[\text{versMed}]}(u)$ consiste en un port à prendre à partir de u pour se déplacer vers m ;
- (3) $\text{LR}_i^{\text{St}[\text{depuisMed}]}(u)$ consiste en un port à prendre à partir de m pour se déplacer vers u ;
- (4) $\text{LR}_i^{\text{St}[\text{Porte}]}(u)$ contient l’identifiant de la fibre contenant u (c.-à-d., l’étiquette de u^\downarrow donnée par un schéma de distance pour l’étoile $\text{St}(m)$).

Les parties 1er et 2nd de l’étiquette de u contiennent des informations similaires, mais elles dépendent de l’appartenance de u à un cône ou à un panneau. Si $F(u^\downarrow)$ est un panneau, alors $\text{LR}_i^{1\text{er}}(u)$ est composée des quatre informations suivantes :

- (1) $\text{LR}_i^{1\text{er}[\text{Emp_LDT}]}(u)$ est l’étiquette de distance d’arbre de la première empreinte u_1 de u sur la bordure $\partial^* F(u^\downarrow)$;
- (2) $\text{LR}_i^{1\text{er}[\text{Emp_LRT}]}(u)$ est l’étiquette de routage de u_1 dans l’arbre $\partial^* F(u^\downarrow)$;
- (3) $\text{LR}_i^{1\text{er}[\text{versEmp}]}(u)$ vaut $\text{port}(u, u_1)$;
- (4) $\text{LR}_i^{1\text{er}[\text{Dist}]}(u)$ est la distance $d_G(u, u_1)$.

La partie 2nd $\text{LR}_i^{2\text{nd}}(u)$ de l’étiquette de u est définie de façon similaire, relativement à la seconde empreinte u_2 de u sur $\partial^* F(u^\downarrow)$.

Si $F(u^\downarrow)$ est un cône, alors il est voisin d’exactly deux panneaux $F(w_1)$ et $F(w_2)$. Les composantes respectives $\text{LR}_i^{1\text{er}}(u)$ et $\text{LR}_i^{2\text{nd}}(u)$ des parties 1er et 2nd de $\text{LR}_i(u)$, consistent toutes deux en quatre informations. $\text{LR}_i^{1\text{er}}(u)$ est, par exemple, composée des données suivantes :

- (1) $\text{LR}_i^{1\text{er}[\text{Porte_LDT}]}(u)$ consiste en l’étiquette de distance d’arbre de la porte u_1^+ de u sur le panneau $F(w_1)$;
- (2) $\text{LR}_i^{1\text{er}[\text{Porte_LRT}]}(u)$ est l’étiquette de routage de u_1^+ sur l’arbre $\partial^* F(w_1)$;
- (3) $\text{LR}_i^{1\text{er}[\text{versPorte}]}(u)$ contient $\text{port}(u, u_1^+)$;
- (4) $\text{LR}_i^{1\text{er}[\text{depuisPorte}]}(u)$ est un port $\text{port}(u_1^+, \text{jum}(u_1^+))$ de u_1^+ vers $\text{jum}(u_1^+)$.

La partie 2nd $\text{LR}_i^{2\text{nd}}(u)$ de l’étiquette de u est définie de façon similaire, relativement à la porte u_2^+ de u sur le panneau $F(w_2)$.

Nous supposons qu'aucun port ne reçoit l'identifiant 0. Si **ROUTAGE** retourne 0, ou si un sommet stocke un port 0, cela signifie qu'aucun déplacement n'est requis. L'algorithme d'encodage est donc le suivant :

Algorithme 3 : **ENC_ROUT** ($G, \text{LR}(V)$)

Entrée : Un graphe médian sans cube $G = (V, E)$ et un étiquetage $\text{LD}(V)$ de ses sommets consistant, initialement, en un identifiant unique $\text{id}(u)$ donné à chaque sommet $u \in V$

- 1 si $V = \{v\}$ alors stop ;
- 2 Trouver un sommet médian m de G ;
- 3 $L_{\text{St}(m)}(\text{St}(m)) \leftarrow \text{Enc_Etoile_Med}(\text{St}(m))$;
- 4 $\mathcal{F}_m \leftarrow \{F(x) : x \in \text{St}(m)\}$;
- 5 **pour chaque** panneau $F(x) \in \mathcal{F}_m$ **faire**
- 6 $\text{LD}_{\partial^* F(x)}(\partial^* F(x)) \leftarrow \text{Enc_Dist_Arbre}(\partial^* F(x))$;
- 7 $\text{LR}_{\partial^* F(x)}(\partial^* F(x)) \leftarrow \text{Enc_Rout_Arbre}(\partial^* F(x))$;
- 8 **pour chaque** $u \in F(x)$ **faire**
- 9 Trouver les deux empreintes u_1 et u_2 de u sur $\partial^* F(x)$;
- 10 $L_{\text{St}} \leftarrow (\text{id}(m), \text{port}(u, m), \text{port}(m, u), L_{\text{St}(m)}(x))$;
- 11 $L_{1\text{er}} \leftarrow (\text{LD}_{\partial^* F(x)}(u), \text{LR}_{\partial^* F(x)}(u), \text{port}(u, u_1), d_G(u, u_1))$;
- 12 $L_{2\text{nd}} \leftarrow (\text{LD}_{\partial^* F(x)}(u), \text{LR}_{\partial^* F(x)}(u), \text{port}(u, u_2), d_G(u, u_2))$;
- 13 $\text{LR}(u) \leftarrow \text{LR}(u) \circ (L_{\text{St}}, L_{1\text{er}}, L_{2\text{nd}})$;
- 14 **ENC_ROUT** ($F(x), \text{LR}(V)$) ;
- 15 **pour chaque** cône $F(x) \in \mathcal{F}_m$ **faire**
- 16 Soient $F(w_1)$ et $F(w_2)$ les panneaux voisins de $F(x)$;
- 17 Supposons $F(w_2)$ le panneau 2nd voisin de $F(x)$;
- 18 $\text{LD}_{\partial^* F(w_1)}(\partial^* F(w_1)) \leftarrow \text{Enc_Dist_Arbre}(\partial^* F(w_1))$;
- 19 $\text{LD}_{\partial^* F(w_2)}(\partial^* F(w_2)) \leftarrow \text{Enc_Dist_Arbre}(\partial^* F(w_2))$;
- 20 $\text{LR}_{\partial^* F(w_1)}(\partial^* F(w_1)) \leftarrow \text{Enc_Rout_Arbre}(\partial^* F(w_1))$;
- 21 $\text{LR}_{\partial^* F(w_2)}(\partial^* F(w_2)) \leftarrow \text{Enc_Rout_Arbre}(\partial^* F(w_2))$;
- 22 **pour chaque** $u \in F(x)$ **faire**
- 23 Trouver la porte u_1^+ de u sur $F(w_1)$;
- 24 Trouver la porte u_2^+ de u sur $F(w_2)$;
- 25 $L_{\text{St}} \leftarrow (\text{id}(m), \text{port}(u, m), \text{port}(m, u), L_{\text{St}(m)}(x))$;
- 26 $L_{1\text{er}} \leftarrow (\text{LD}_{\partial^* F(w_1)}(u), \text{LR}_{\partial^* F(w_1)}(u), \text{port}(u, u_1^+), \text{port}(u_1^+, \text{jum}(u_1^+)))$;
- 27 $L_{2\text{nd}} \leftarrow (\text{LD}_{\partial^* F(w_2)}(u), \text{LR}_{\partial^* F(w_2)}(u), \text{port}(u, u_2^+), \text{port}(u_2^+, \text{jum}(u_2^+)))$;
- 28 $\text{LR}(u) \leftarrow \text{LR}(u) \circ (L_{\text{St}}, L_{1\text{er}}, L_{2\text{nd}})$;
- 29 **ENC_ROUT** ($F(x), \text{LR}(V)$) ;

7.8.3 Requêtes de routage

Soient u et v deux sommets arbitraires d'un graphe médian sans cube G , et soient $\text{LR}(u)$ et $\text{LR}(v)$ les étiquettes respectives leur ayant été attribuées par **ENC_ROUT**. Nous décrivons ici comment l'algorithme **ROUTAGE** peut décider quel port de u prendre pour envoyer un message de u vers v .

7.8.3.1 L'algorithme

Commençons par la description formelle de l'algorithme de routage **ROUTAGE**. Les fonctions utilisées par cet algorithme pour router d'un panneau vers un cône, d'un cône vers un panneau, d'un cône vers un cône ou entre deux sommets éloignés seront données dans une section (7.8.3.2) suivante.

Algorithme 4 : ROUTAGE ($LR(u), LR(v)$)

Entrée : Les étiquettes $LR(u)$ et $LR(v)$ de deux sommets u et v de G , où u est la source et v la cible

Sortie : $\text{port}(u, v)$

- 1 si $LR_0(u) = LR_0(v)$ /* $u = v$ */ alors retourner 0 ;
- 2 Soit i le plus grand entier tel que $LR_i^{\text{St[Med]}}(u) = LR_i^{\text{St[Med]}}(v)$;
- 3 $d \leftarrow \text{Dist_Etoile}(LR_i^{\text{St[Porte]}}(u), LR_i^{\text{St[Porte]}}(v))$; // $d_G(u^\downarrow, v^\downarrow)$
- 4 $d_u \leftarrow \text{Dist_Etoile}(LR_i^{\text{St[Porte]}}(u), 0)$; // $d_G(u^\downarrow, m)$
- 5 $d_v \leftarrow \text{Dist_Etoile}(LR_i^{\text{St[Porte]}}(v), 0)$; // $d_G(v^\downarrow, m)$
- 6 si $d = 1$ et $d_u = 1$ alors retourner $\text{Rout_PvC}(LR_i(u), LR_i(v))$;
- 7 si $d = 1$ et $d_v = 1$ alors retourner $\text{Rout_CvP}(LR_i(u), LR_i(v))$;
- 8 si $d = 2$ et $d_u = d_v = 2$ alors retourner $\text{Rout_CvC}(LR_i(u), LR_i(v))$;
- 9 retourner $\text{Rout_Eloignes}(LR_i(u), LR_i(v), LR_0(u))$;

7.8.3.2 Description et fonctions

De même que pour les requêtes de distance, la fonction de décodage pour le routage commence par déterminer si u et v sont 1-voisins, 2-voisins ou éloignés ainsi que le type (cône ou panneau) des fibres contenant ces deux sommets. Ce calcul est fait comme expliqué en section 7.7.2. Encore une fois, supposons que i soit le numéro de la première étape récursive de l'encodage pour laquelle u et v ne sont plus proches. Dénotons par m le médian du sous-graphe contenant u et v considéré à cette étape. Dénotons aussi par $F(x)$ et $F(y)$ les fibres respectives contenant u et v . Rappelons qu'alors x dénote la porte de u sur $\text{St}(m)$ et y dénote la celle de v .

Si u et v sont 1-voisins, la réponse à la requête est calculée différemment selon que la source u soit dans un cône ou un panneau. Si u appartient à un cône $F(x)$ (et donc v est dans un panneau $F(y)$), nous utilisons la fonction Rout_CvP . Cette fonction détermine quelle partie ($LR_i^{\text{1er}}(v)$ ou $LR_i^{\text{2nd}}(v)$) de $LR_i(v)$ contient les informations relatives à la porte u^+ de u sur $F(y)$. Puis elle retourne le port $\text{port}(u, u^+)$, stocké en tant que $LR_i^{\text{1er[versPorte]}}(u)$ ou $LR_i^{\text{2nd[versPorte]}}(u)$ dans l'étiquette de u .

```

fonction Rout_CvP(LRi(u), LRi(v)):
  dir ← 2nd ; // remarque 24
  si LRiSt[Porte](v) = min{j : j ∈ LRiSt[Porte](u)} alors
    dir ← 1er ;
  retourner LRidir[versPorte](u) ;

```

Si $F(x)$ est un panneau (et $F(y)$ un cône), alors u a stocké sa distance vers ses deux empreintes u_1 et u_2 sur la bordure $\partial^*F(x)$ et v a stocké sa distance vers sa porte v^+ sur $F(x)$ (v^+ appartenant aussi à $\partial^*F(x)$) Quand u ne coïncide pas avec v^+ , la fonction Rout_PvC décode des distances dans les arbres pour calculer $\min\{d_G(u, u_1) + d_T(u_1, v^+), d_G(u, u_2) + d_T(u_2, v^+)\}$, et retourne le port vers l'empreinte de u atteignant ce minimum. Si u appartient à la bordure de $F(x)$, alors deux cas sont à distinguer. Si $u = v^+$, l'algorithme peut retrouver dans $LR_i(v)$ le port de $\text{jum}(v^+)$ vers $v^+ = u$ et le retourner. Si $u \neq v^+$, comme $LR_i(u)$ et $LR_i(v)$ contiennent des étiquettes de u et v^+ données par un schéma de routage dans l'arbre $\partial^*F(x)$, Rout_PvC calcule $\text{port}(u, v^+)$ en utilisant le décodeur de routage dans les arbres et retourne ce port.

```

fonction Rout_PvC(LRi(u), LRi(v)):
  dirv ← 2nd ;
  si LRiSt[Porte](u) = min{j : j ∈ LRiSt[Porte](v)} alors
    dirv ← 1er ;

  // Si u est sur une frontière
  si LRi1er[versPorte](u) = 0 ou LRi2nd[versPorte](u) = 0 alors
    Soit diru ∈ {1er, 2nd} tel que LRidiru[versPorte](u) = 0 ;
    si LRidiru[Porte_LDT](u) = LRidirv[Porte_LDT](v) alors
      // u est la porte de v sur le panneau F(x)
      retourner LRidirv[depuisPorte](v) ;
    retourner Rout_Arbre(LRidiru[Porte_LRT](u), LRidirv[Porte_LRT](v)) ;

  d1er ← LRi1er[Dist](u) + Dist_Arbre(LRi1er[Porte_LDT](u), LRidirv[Porte_LDT](v)) ;
  d2nd ← LRi2nd[Dist](u) + Dist_Arbre(LRi2nd[Porte_LDT](u), LRidirv[Porte_LDT](v)) ;
  diru ← 1er ;
  si d2nd = min{d1er, d2nd} alors
    // u2nd est la porte de u sur un plus court chemin vers v
    diru ← 2nd ;
  retourner LRidiru[versPorte](u) ;

```

Lorsque u et v sont 2-voisins, $F(x)$ et $F(y)$ sont deux cônes et la fonction Rout_CvC est alors similaire à la fonction Rout_CvP. Le panneau $F(w)$ voisin de $F(x)$ et $F(y)$ peut être trouvé en inspectant $LR_i^{\text{St[Porte]}}(u)$ et $LR_i^{\text{St[Porte]}}(v)$. Comme dans le cas de Rout_CvP, la fonction Rout_2-Voisins retourne le porte $\text{port}(u, u^+)$ de u vers sa porte u^+ sur $F(w)$.

```

fonction Rout_CvC( $LR_i(u)$ ,  $LR_i(v)$ ):
  dir  $\leftarrow$  2nd ;
  si  $LR_i^{\text{St[Porte]}}(u) \cap LR_i^{\text{St[Porte]}}(v) = \min\{j : j \in LR_i^{\text{St[Porte]}}(u)\}$  alors
    dir  $\leftarrow$  1er ;
  retourner  $LR_i^{\text{dir[versPorte]}}(u)$  ;

```

Enfin, si u et v sont éloignés, deux cas doivent être considérés selon que u coïncide ou non avec le médian m . Si $u \neq m$, alors $LR_i(u)$ contient $\text{port}(u, m)$. Et comme il existe un plus court chemin entre u et v passant par m , `Rout_Eloignes` retourne $\text{port}(u, m)$. Si u coïncide avec m , alors le port $\text{port}(m, v)$ n'a pas pu être stocké dans $LR_i(u)$, mais il est présent dans $LR_i^{\text{St[depuisMed]}}(v)$ et c'est ce que `Rout_Eloignes` retourne.

```

fonction Rout_Eloignes( $LR_i(u)$ ,  $LR_i(v)$ ,  $id(u)$ ):
  si  $LR_i^{\text{St[Med]}}(v) = id(u)$  alors
    retourner  $LR_i^{\text{St[depuisMed]}}(v)$ 
  retourner  $LR_i^{\text{St[versMed]}}(u)$  ;

```

Remarque 25. Des analyses semblables à celles de `ENC_DIST` et `DISTANCE` peuvent être appliquées à `ENC_ROUT` et `ROUTAGE` pour montrer le théorème 21.

Chapitre 8

Graphes pontés sans K_4

Sommaire

8.1	Résultat principal	148
8.2	L'idée du schéma	149
8.3	Propriétés	150
8.4	Étoiles dans les graphes pontés sans K_4	155
8.5	Fibres des graphes pontés sans K_4	156
	8.5.1 Frontières et bordures	161
	8.5.2 Projections et représentants	162
8.6	Classification des sommets et plus courts chemins	166
8.7	Schéma de distance 4-approché pour les graphes pontés sans K_4	172
	8.7.1 Encodage	172
	8.7.2 Requêtes de distances	174
	8.7.3 Correction et complexité	176

Dans ce chapitre, nous décrivons un schéma d'étiquetage de distance approché pour les graphes pontés sans K_4 . Ce schéma utilise des étiquettes de longueur poly-logarithmique et permet un décodage des distances avec une erreur multiplicative de 4 dans le pire des cas.

8.1 Résultat principal

À moins que le contraire ne soit précisé, les graphes $G = (V, E)$ étudiés dans ce chapitre sont pontés et n'admettent pas K_4 pour sous-graphe. Dans ce qui suit, pour un $k \in \mathbb{N}^*$, nous appelons *schéma de distance k -approché*, un schéma d'étiquetage de distance dans lequel la fonction de décodage retourne une valeur entre la distance exacte et k fois celle-ci. Pour ces graphes, nous montrons le théorème suivant :

Théorème 22. *La classe \mathcal{G} des graphes pontés, sans K_4 , avec n sommets, admet un schéma d'étiquetage de distance 4-approché utilisant des étiquettes de $O(\log^3 n)$ bits, attribuées en temps polynomial et décodées en temps constant $O(1)$.*

Dans la section 8.2 suivante, nous présentons les similarités et les différences notoires entre le schéma de distance pour les graphes médians sans cube proposé dans le chapitre 7 et le schéma que nous décrivons dans ce chapitre-ci. Nous donnons ensuite, en section 8.3, quelques propriétés structurelles des graphes pontés sans K_4 . Dans les sections 8.4 et 8.5, nous introduisons respectivement les notions d'étoiles et de fibres pour les graphes pontés sans K_4 , puis nous donnons les propriétés de ces objets utiles à notre schéma. Dans la section 8.6, nous définissons une classification de paires de sommets et précisons comment calculer des distances approchées entre eux. Plus précisément, nous décrivons les paires de sommets pour lesquelles nous ne ferons pas d'erreur sur la distance, celles sur lesquelles nous donnerons une distance entre l'exacte et deux fois l'exacte, et celles pour lesquelles nous ferons une erreur d'un facteur multiplicatif inférieure ou égal à 4. Enfin, dans la section 8.7, nous présentons notre schéma de distance 4-approché pour les graphes pontés sans K_4 .

8.2 L'idée du schéma

Il n'est pas nécessaire que nous rappelions ici l'idée générale de notre schéma pour les graphes pontés sans K_4 dans la mesure où les grandes lignes sont les mêmes que celles du chapitre 7 précédent. À savoir, le schéma est basé sur une partition récursive de nos graphes en une étoile et ses fibres : des cônes et des panneaux. En revanche, la structure et les propriétés de ces étoiles et ces fibres sont très différentes de celles que nous avons dans les graphes médians. Dans cette section, nous présentons donc rapidement ces différences, et nous essayons de donner l'intuition des difficultés techniques qu'elles induisent.

Tout d'abord, une étoile $\text{St}(z)$ dans un graphe ponté $G = (V, E)$ sans K_4 n'est ni convexe, ni portée. Ceci dit, il est tout de même possible de définir des fibres partant des sommets de cette étoile, de sorte qu'un sommet soit toujours dans une et une seule fibre (c.-à-d., une étoile et ses fibres partitionnent le graphe). Ensuite, les fibres $F(x)$ de $\text{St}(z)$ ne sont pas convexes. Elles sont, en revanche, connexes et isométriques, et induisent (par conséquent) des graphes pontés (puisque, étant isométriques, elles ne peuvent pas contenir de cycles isométriques plus longs que le graphe initial). Sans cela, nous ne pourrions pas obtenir un facteur multiplicatif constant puisque les erreurs seraient accumulées à chaque étape récursive. De plus, ces fibres ont la propriété intéressante d'être étoilées (c.-à-d., elle contiennent l'intervalle $I(u, x)$ pour tout $u \in F(x)$). De même que dans les graphes médians, nous étudierons la bordure de ces fibres

(définie comme précédemment). Nous verrons que cette dernière n'est pas tout à fait un arbre, mais un "arbre étoilé" (que nous définirons plus loin). Malheureusement, ces bordures ne sont pas non plus isométriques. Cela explique, en partie, que nous obtenions un schéma de distance approché. En effet, nous verrons que les distances dans les bordures peuvent être le double des distances du graphe. Toujours concernant les frontières et bordures, elles ne sont plus portées (ou quasi-portées) dans le cas des graphes pontés (sans K_4). Nous verrons toutefois qu'un sommet u dans un panneau $F(x)$ admet deux "quasi-représentants" sur la bordure de $F(x)$, et qu'un sommet v dans un cône $F(y)$ admet un "représentant" sur chaque frontière de $F(y)$ (et il n'y en a que deux). Ces "représentants" et "quasi-représentants" auront la garantie d'être sur un chemin de longueur inférieure à quatre fois celle d'un plus court chemin.

Avec ces propriétés en tête, les algorithmes d'encodage et décodage seront finalement assez proches de ceux exposés dans le chapitre 7. En fait, en autorisant une erreur multiplicative de 4, nous verrons que dans presque tous les cas, la distance retournée entre deux sommets u et v sera celle passant par le centre m de l'étoile $\text{St}(m)$ de la partition. Les lemmes 52 et 53 indiquent respectivement quand cette valeur correspond à une distance exacte et quand elle donne une distance approchée (à facteur 2). Le cas u et v 1pc-voisins, celui nécessitant un calcul plus technique et conduisant à une erreur multiplicative de 4 dans le pire des cas, sera défini et présenté plus tard.

8.3 Propriétés

Dans cette section, nous commençons par décrire la structure des triangles métriques et des intervalles dans les graphes pontés sans K_4 . Nous introduisons ensuite la notions d'ensemble étoilé et décrivons la structure de l'intersection entre un intervalle et un arbre étoilé. Rappelons que, lorsque nous parlons d'un triangle métrique, c'est souvent son enveloppe convexe que nous considérons.

Le lemme 38 présente une propriété des graphes pontés sans K_4 utile à la preuve du lemme 39 qui suivra. Pour l'énoncer, nous donnons le nom de *triangle plat* au graphe $T_k = (V_k, E_k)$ défini par

- $V_k := \{u_i^j : 0 \leq j \leq k \text{ et } 0 \leq i \leq j\}$;
- $E_k := \{u_i^j u_{i'}^{j'} : u_i^j, u_{i'}^{j'} \in V_k, (j = j' \text{ et } i' = i - 1) \text{ ou } (j' = j - 1 \text{ et } i - 1 \leq i' \leq i)\}$.

Un motif correspondant à un triangle plat est donné en figure 8.1 (gauche).

Lemme 38. *L'enveloppe convexe d'un triangle métrique (u, v, w) dans un graphe ponté sans K_4 contient toujours un triangle plat en tant que sous-graphe induit.*

Démonstration. Posons $k := d_G(v, w)$ et considérons un plus court chemin P entre

v et w . Nous renommons ses sommets $v =: u_0^k, u_1^k, \dots, u_k^k := w$, où u_i^k est le sommet à distance i de $v = u_0^k$ sur P . D'après le théorème 10, chaque sommet de P est aussi à distance k de u . Pour tout $i \in \{0, \dots, k-1\}$, nous pouvons appliquer la condition du triangle sur u_i^k, u_{i+1}^k et $u_0^0 := u$ pour dériver un sommet $u_i^{k-1} \sim u_i^k, u_{i+1}^k$ à distance $k-1$ de u_0^0 . Nous voulons alors montrer que :

- (1) $\forall i, i' \in \{0, \dots, k-1\}$, avec $i < i'$, $u_i^{k-1} \neq u_{i'}^{k-1}$;
- (2) $\forall j \in \{0, \dots, k-2\}$, $u_j^{k-1} \sim u_{j+1}^{k-1}$.

Pour montrer (1), faisons l'hypothèse absurde qu'il existe $i \neq i'$ tels que $u_i^{k-1} = u_{i'}^{k-1} =: x$. Alors x est adjacent à $u_i^k, u_{i+1}^k, u_{i'}^k$ et $u_{i'+1}^k$. Rappelons que chaque sommet de $I(u_0^k, u_k^k)$ est à distance k de u_0^0 . En particulier, chaque sommet de $I(u_i^k, u_{i'+1}^k)$ doit être à distance k de u_0^0 . Par conséquent, $u_i^{k-1} \notin I(u_i^k, u_{i'+1}^k)$. Il s'en suit que $\{u_i^k, u_{i+1}^k, u_{i'}^k, u_{i'+1}^k\}$ induit une clique, et que P n'est pas un plus court chemin. La propriété (2) découle de la convexité des boules $B_{k-1}(u_0^0)$. En effet, nous avons : $u_j^{k-1} \neq u_{j+1}^{k-1}$; $u_j^{k-1}, u_{j+1}^{k-1} \sim u_j^k$; et $u_j^k \notin B_{k-1}(u_0^0)$. Nous pouvons en conclure que $u_j^{k-1} \sim u_{j+1}^{k-1}$.

Nous pouvons alors remplacer P , par le chemin P' de longueur $k-1$ induit par $\{u_0^{k-1}, \dots, u_{k-1}^{k-1}\}$ et lui appliquer les mêmes arguments. Ainsi de suite jusqu'à dériver $\{u_0^0\}$.

Le triangle métrique (u, v, w) contient donc un triangle plat en tant que sous-graphe. Il nous reste à prouver que ce sous-graphe est maximal en nombre d'arêtes (c.-à-d., qu'il est induit). Pour cela, il suffit de montrer que pour tout $i \in 1, \dots, k-1$, les seuls voisins de u_i^k dans le triangle plat (généré par P et u_0^0) sont $u_{i-1}^k, u_{i+1}^k, u_{i-1}^{k-1}$ et u_i^{k-1} . De même, les seuls voisins de u_0^k dans ce triangle sont u_1^k et u_0^{k-1} , et ceux de u_k^k sont u_{k-1}^k et u_{k-1}^{k-1} . En effet, si un sommet u_i^k possède un autre voisin dans P que ceux listés précédemment, alors P n'est plus un plus court chemin. Concernant les sommets sur le niveau $k-1$, u_i^k ne peut pas avoir plus de deux voisins (u_{i-1}^{k-1} et u_i^{k-1}) de par la convexité de la boule $B_{k-1}(u_0^0)$ à laquelle n'appartient pas u_i^k (c.-à-d., ajouter un troisième sommet adjacent à u_i^k dans $I(u_i^k, u_0^0)$ créerait un K_4). Enfin, tout lien entre u_i^k et un sommet $u_{i'}^{k-1}$ avec $k' < k-1$ et $i' \in \{0, \dots, k'\}$ serait contradictoire avec le théorème 10. \square

Le lemme qui suit nous sera utile dans les preuves qui font intervenir des sommets dans des triangles métriques. Nous nous réfèrerons directement à la structure de ces triangles, décrite dans ce lemme 39, pour exhiber des sommets où des liens entre sommets dans un triangle métrique.

Lemme 39. *L'enveloppe convexe d'un triangle métrique (u, v, w) dans un graphe ponté $G = (V, E)$ sans K_4 est exactement un triangle plat.*

Démonstration. Il nous suffit ici de prouver que le triangle plat T induit dans (u, v, w) par le lemme 38 est maximal en nombre de sommets. Nous montrons

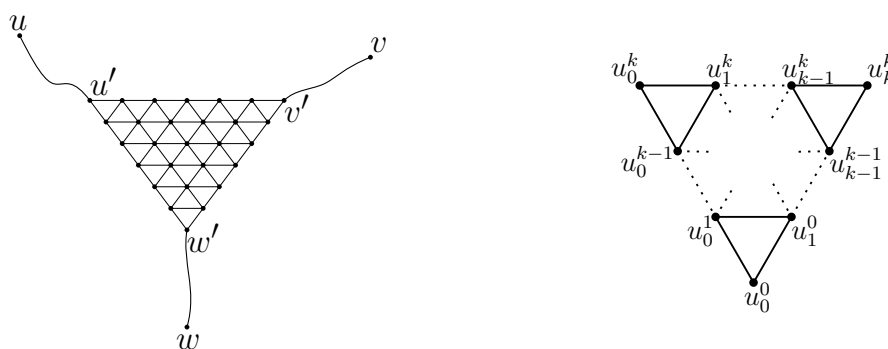
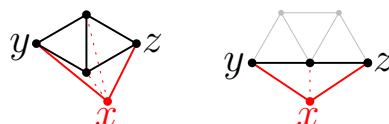


FIGURE 8.1 – À gauche, l’illustration d’un motif présent dans chaque quasi-médian (u', v', w') d’un triplet (u, v, w) de sommets d’un graphe ponté sans K_4 . À droite, une aide à la preuve du lemme 38.

pour cela qu’il est localement convexe (et donc convexe). Supposons, par l’absurde, qu’il existe un sommet x dans (u, v, w) n’appartenant pas au triangle plat illustré T en 8.1. Un quasi-médian étant nécessairement connexe, x doit être connecté à un sommet y de T . Supposons que cette connexion soit une arête, (sinon nous remplaçons x par le sommet x' sur un (x, y) -chemin tel que $x' \sim y$). Puisque nous avons supposé, par l’absurde, que T n’est pas localement convexe, x doit appartenir à l’intervalle $I(y, z)$ pour un certain z à distance 2 de y . L’intervalle entre y et z dans T peut être des deux formes illustrées ci-dessous (en noir l’intervalle $I_T(y, z)$, en rouge la présence de x dans $I_G(y, z)$, en gris des sommets issus de la structure du triangle plat) :



Dans le dessin de gauche, la convexité de $N_G[y]$ implique que x soit adjacent aux deux voisins de y dans $I_T(y, z)$, faisant apparaître un K_4 . Dans celui de droite, x doit être relié à l’unique voisin de y dans $I_T(y, z)$ pour ne pas créer de C_4 isométrique, mais cela crée alors un W_5 (et donc un long cycle isométrique, ce qui contredit que G est ponté). \square

Soient u et v deux sommets de la grille triangulaire “complète” (c.-à-d., le squelette du pavage du plan euclidien par des triangles équilatéraux de côté 1). Une *grille triangulaire rongée* est obtenue en retirant (ou non) itérativement des sommets de degré 3 dans le “parallélogramme” induit par $I(u, v)$ (un exemple est donné dans la figure 8.2). Initialement, dans la grille triangulaire complète, $I(u, v)$ contient deux sommets de degré 2 (dans $I(u, v)$) et deux sommets de degré 3. Ces sommets sont appelés des *coins convexes*. Les sommets sur les plus courts chemins reliant ces coins externes sont des *bords*. Les sommets restant

sont des *sommets internes*. Un sommet w de degré 3 dans $I(u, v)$ possède un voisin $w_u \in W(u, v)$, un voisin $w_v \in W(v, u)$ et un voisin $w' \in W^=(u, v)$. Si w est supprimé, w_u et w_v deviennent des coins convexes et w' devient un coin concave. Par la suite, lors de la suppression d'un de ses voisins, un sommet interne devient un coin concave, un coin concave devient un bord, un bord devient un coin convexe et un coin convexe devient un bord (ces évolutions de statuts sont indiquées dans la table 8.1). Remarquons que le voisinage d'un sommet w retiré est toujours un P_3 induit dans $I(u, v)$ (c.-à-d., un P_3 convexe de $I(u, v) \setminus \{w\}$).

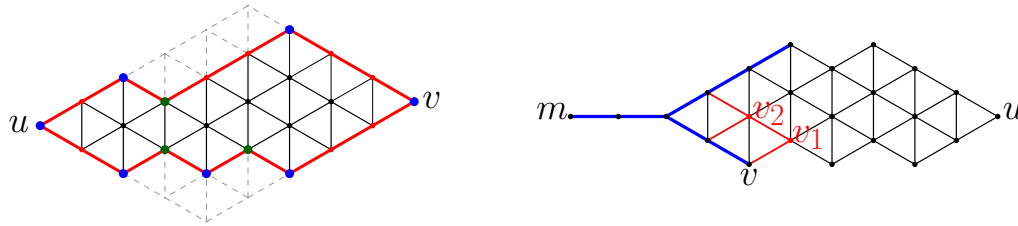
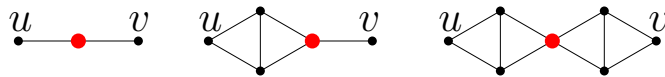


FIGURE 8.2 – À gauche, un exemple de grille triangulaire rongée (induite par l'intervalle entre u et v , voir lemme 40). Les coins concaves sont en vert, les convexes sont en bleu. Les bords sont en rouge et les sommets internes en noir. À droite, une autre grille triangulaire rongée pour illustrer le lemme 41.

Remarque 26. Les trois sommets rouges des dessins ci-dessous sont, en quelque sorte, des cas “dégénérés”. Notons que ce sont, de gauche à droite, un bord, un coin convexe et un bord. De plus, le sommet u de la première figure et les sommets v de la première et de la deuxième figure sont aussi des bords.



Voisins retirés	bord	concave	convexe	interne
1	convexe	bord	bord	concave
2	bord	convexe	convexe	bord
3	convexe	bord	bord	convexe
4	bord	convexe	convexe	bord

TABLE 8.1 – Évolution du statut d'un sommet d'une grille rongée en fonction du nombre de voisins qui lui sont retirés.

Lemme 40. *L'intervalle $I(u, v)$ entre deux sommets u et v d'un graphe ponté $G = (V, E)$ sans K_4 est une grille triangulaire rongée.*

Démonstration. Soient u et v deux sommets de G et soit $P^0 \subseteq I(u, v)$ un plus court chemin entre u et v dans G (c.-à-d., P^0 est une grille triangulaire rongée

minimale, ainsi qu'un potentiel intervalle minimal). Alors, soit $P^0 = I(u, v)$, soit pour tout $x_1 \in \{a \in N_G(b) : b \in P^0\}$ tel que $P^1 := P^0 \cup \{x_1\} \subseteq I(u, v)$, P^1 est une grille triangulaire rongée (c.-à-d., x_1 est connecté aux sommets d'un P_3 induit de P^0). En effet, supposons que $P^0 \subsetneq I(u, v)$ (sinon nous avons terminé). Soit x_1 un sommet tel que $P^1 \subseteq I(u, v)$. Si x_1 n'est connecté qu'à un unique sommet ou une unique arête de P^0 , alors il ne peut clairement pas être sur un plus court chemin. Si x_1 est connecté à exactement deux sommets non-adjacents de P^0 ou a plus de 3 sommets, alors il crée un raccourci (et contredit que P^0 soit un plus court chemin). Donc x est connecté à exactement 3 sommets de P^0 , et si ces trois sommets sont non consécutifs (c.-à-d., pas dans un P_3), alors x_1 crée à nouveau un raccourci.

Supposons maintenant que, pour un certain $k \geq 1$, $|I(u, v)| \geq |P^0| + k - 1$ et que nous ayons pu itérativement ajouter des sommets x_1, \dots, x_{k-1} à P^0 pour obtenir une grille triangulaire rongée P^{k-1} . Alors, soit $P^{k-1} = I(u, v)$ (nous supposons que ce n'est pas le cas), soit pour tout x_k tel que $P^k := P^{k-1} \cup \{x_k\} \subseteq I(u, v)$, P^k est une grille triangulaire rongée (c.-à-d., x_k est relié à chaque sommet d'un P_3 induit dans P^{k-1} , et aucun des sommets du P_3 n'est interne, autrement dit, P_3 est convexe). En effet, si x_k n'est relié qu'à un ou deux sommets de P^{k-1} , alors il ne peut pas être dans $I(u, v)$. Il en est de même si x_k est relié à plus de trois sommets. Donc x_k est relié à trois sommets. Si ces trois sommets sont un P_3 non induit, alors x_k et son voisinage forment un K_4 , et si ce n'est pas un P_3 , alors x_k crée un raccourci. Enfin, si le voisinage de x_k est un P_3 avec un ou deux sommets internes, alors G contient un W_4 , et si les trois sommets sont internes, G contient un W_5 . Ainsi, x_k est forcément connecté à un P_3 du bord de P^{k-1} , et P^k est donc une grille triangulaire rongée. \square

Remarque 27. L'enveloppe convexe des triangles métriques et les intervalles d'un graphe ponté sans K_4 induisent des graphes planaires. Pour autant, les graphes pontés sans K_4 contiennent tous les sous-graphes en tant que mineurs et sont donc, en particulier, non-planaires. En revanche, les graphes pontés $G = (V, E)$ sans K_4 étant 3-dégénérés, leur nombre d'arêtes est d'au plus $O(|V|)$ (et leur densité est constante).

Un chemin P d'un arbre T (enraciné en r) est dit *montant* si P est entièrement inclus dans une branche de T (c.-à-d., $\forall u, v \in P$, soit $I_T(u, r) \subseteq I_T(v, r)$, soit $I_T(v, r) \subseteq I_T(u, r)$). Un sous-ensemble S de sommets d'un graphe quelconque G est dit *étoilé* (par rapport à $s \in S$) si, pour tout $s' \in S$, $I(s, s') \subseteq S$. Cet ensemble S est appelé un *arbre étoilé* (enraciné en $s \in S$) si, pour tout $s' \in S$, $I(s, s')$ est un chemin.

Lemme 41. Soit T un arbre étoilé (enraciné en m) dans un graphe ponté $G = (V, E)$ sans K_4 , et soit $u \in V \setminus T$. Alors $I(u, m) \cap T$ est inclus dans les deux chemins montant de m jusqu'aux premiers coins convexes non dégénérés (ces deux coins sont dits extrémaux par rapport à m).

Démonstration. Posons $T' := I(u, m) \cap T$. Puisque, en particulier, $T' \subseteq T$, T' est une forêt. Considérons maintenant deux sommets v et w de T' . Alors $v \in I(u, m)$ implique $I(v, m) \subseteq I(u, m)$. Puisque $v \in T$ et T est étoilé, nous en déduisons que $I(v, m) \subseteq T$. Nous en concluons que $I(v, m) \subseteq T'$ et, de la même façon, que $I(w, m) \subseteq T'$. T' est donc un arbre étoilé.

De par la structure de $I(u, m)$ décrite dans le lemme 40, chaque sommet de T' possède au plus deux fils. Nous voulons montrer que, pour tout $v \in T' \setminus \{m\}$, soit v est une feuille, soit il possède un unique fils étant un bord ou un coin convexe. De plus, si v est un coin convexe non-dégénéré, alors c'est une feuille de T' . Supposons, par l'absurde, que ce ne soit pas le cas. Les deux fils potentiels de m sont des bords ou des coins convexes, choisissons donc le sommet v le plus proche de m dans T' ayant pour fils un sommet interne ou un coin concave v_1 . Remarquons que v est un bord ou un coin convexe (autrement nous aurions déjà choisi l'un de ses ancêtres pour contre-exemple). En utilisant à nouveau la structure de $I(u, m)$, nous savons qu'il existe un sommet v_2 avec $v_2 \neq v$ et $v_2 \in I(v_1, m)$. Mais alors, si v_2 appartenait à T' , nous aurions choisi le père de v_2 à la place de v , et si v_2 n'appartient pas à T' , cela contredit le fait que T' soit étoilé (voir l'illustration 8.2). \square

8.4 Étoiles dans les graphes pontés sans K_4

Soit $z \in V$ un sommet quelconque de G . Rappelons que, les graphes pontés étant exactement les graphes dans lesquels le voisinage des convexes est convexe, $N_G[z]$ doit être convexe.

Nous pouvons remarquer que, pour tout choix de $u \in V \setminus N[z]$, z ne peut pas appartenir à la projection métrique $\text{Pr}(u, N[z])$. En effet, z doit posséder un voisin z' sur un plus court (u, z) -chemin. Or ce z' est plus proche de u que z , et il appartient à $N[z]$.

Lemme 42. *Soit $G = (V, E)$ un graphe ponté sans K_4 et soit $z \in V$. Alors pour tout $u \in V \setminus \{z\}$, $\text{Pr}(u, N[z])$ consiste soit en un unique sommet, soit en deux sommets reliés par une arête.*

Démonstration. Soient $x, y \in \text{St}(z)$ deux projections (métriques) de u telles que $x \not\sim y$ (c.-à-d., $d_G(x, y) = 2$). Posons $k := d_G(u, x) = d_G(u, y)$. Alors $d_G(u, z) = k + 1$. Par la condition du quadrangle, il doit exister un sommet $u' \sim x, y$ à distance $k - 1$ de u . Puisque G ne peut pas contenir de C_4 induits, (x, y, z, u') doit contenir une corde : xy ou zu' . Puisque $x \not\sim y$, cette corde doit être zu' . Mais alors $u' \in N[z]$ est plus proche de u que x et y , contredisant $x, y \in \text{Pr}(u, N[z])$. Nous avons donc montré que deux projections doivent être reliées par une arête. Si u possédait une troisième projection métrique w , cette dernière devrait donc être reliée aux deux autres. Mais alors w, x, y, z induiraient un K_4 . \square

Soit $u \in V$ un sommet avec deux projections (métriques) x et y sur $N[z]$. Alors, par le lemme 42 et par la condition du triangle, il existe un sommet u' à distance $d_G(x, u) - 1$ de u , et relié à x et à y . De plus, $I(u', z) = \{u', z, x, y\}$. L'étoile $\text{St}(z)$ d'un sommet $z \in V$ correspond à son voisinage $N[z]$ ainsi qu'à tous les triangles que l'on peut dériver (par la condition du triangle) avec les projections sur $N[z]$ étant des arêtes. La figure 8.3 montre un graphe obtenu comme étoile d'un sommet dans un graphe ponté sans K_4 .

Remarque 28. L'étoile d'un sommet dans un graphe ponté peut avoir une structure assez complexe puisque les voisinages ouverts des sommets des graphes pontés sans K_4 correspondent déjà exactement à l'ensemble des graphes de maille au moins 6.

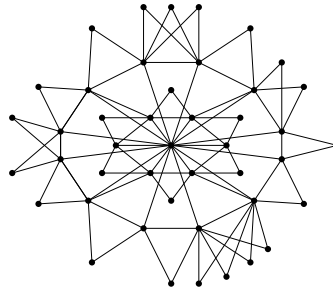


FIGURE 8.3 – Exemple d'une étoile d'un graphe ponté sans K_4 .

8.5 Fibres des graphes pontés sans K_4

Soit z un sommet de G et soit x un sommet de $\text{St}(z)$. Si $x \in N[z]$, nous définissons la *fibre* $F(x)$ de x respectivement à $\text{St}(z)$ comme l'ensemble de tous les sommets de G ayant x pour unique projection sur $N[z]$. Sinon (c.-à-d., si $d_G(x, z) = 2$), $F(x)$ dénote l'ensemble des sommets ayant une arête pour projection sur $\text{St}(z)$ et qui sont plus proche de x que des extrémités de cette arête.

Proposition 13. $\mathcal{F}_z := \{F(x) : x \in \text{St}(z)\}$ est une partition de G .

Pour la suite de cette section, nous considérons des fibres respectivement à $\text{St}(z)$. Une fibre $F(x)$ est un *panneau* si $d_G(x, z) = 1$, un *cône* si $d_G(x, z) = 2$. Deux fibres $F(x)$ et $F(y)$ sont *k-voisines* si $d_{\text{St}(z) \setminus \{z\}}(x, y) = k$.

Nous continuons avec deux lemmes “utilitaires” pour nos preuves futures.

Lemme 43. Soient $F(x)$ et $F(y)$ deux fibres, et soient $u \in F(x)$ et $v \in F(y)$ deux sommets adjacents. Posons $k := d_G(v, y)$. Si $F(y)$ est un cône et $F(x)$ est un panneau, alors $d_G(u, y) = d_G(u, x) \in \{k, k + 1\}$ et $x \sim y$.

Démonstration. Dénotons par y_1 et y_2 les deux voisins de y dans $\text{St}(z)$. Commençons par remarquer que, d'une part, $d_G(v, y) + 1 \geq d_G(v, x)$ et $d_G(v, y) = k$ impliquent que $d_G(v, x) \geq k + 1$ et $d_G(u, x) \geq k$. D'autre part, $d_G(u, x) \leq d_G(u, y) \leq k + 1$ car $d_G(u, x) > d_G(u, y)$ impliquerait que y_1 et y_2 soient deux projections de u sur $N_G(z)$. Par conséquent, $d_G(u, x) \in \{k, k + 1\}$. Il nous reste à prouver que $x \sim y$ et que $d_G(u, x) = d_G(u, y)$ dans les deux cas $d_G(u, x) = k$ et $d_G(u, x) = k + 1$.

Considérons, dans un premier temps, que $d_G(u, x) = k$. Alors $d_G(v, x) = k + 1$ et x doit donc coïncider avec y_1 ou y_2 (sinon v aurait plus de deux projections sur $N_G(z)$). Autrement dit, $x \sim y$. Pour voir que $d_G(u, x) = d_G(u, y)$, supposons par l'absurde que ce ne soit pas le cas (c.-à-d., $d_G(u, y) = k + 1$), et considérons un voisin u' de u tel que $d_G(u', x) = d_G(u, x) - 1$. Alors v et u' appartiennent tous deux à $B_k(y)$, ce qui n'est pas le cas de $u \sim v, u'$. Puisque $B_k(y)$ est convexe, nous déduisons que $v \sim u'$. De ce fait, nous avons $d_G(v, x) = 1 + d_G(u', x) = k$, ce qui est impossible.

Considérons maintenant le cas dans lequel $d_G(u, x) = k + 1$. Puisque $d_G(u, x) \leq d_G(u, y) \leq k + 1$, nous en déduisons que $d_G(u, x) = d_G(u, y) = k + 1$. Nous avons aussi, $d_G(u, z) = k + 2 = d_G(u, y_1)$. Par la condition du triangle, il doit donc exister un sommet $w \sim z, y_1$ à distance $k + 1$ de u . w est nécessairement égal à x , autrement u admettrait deux projections sur $\text{St}(z)$ et ne serait donc pas dans un panneau. Mais $w = x \sim y_1$ implique que $x = y_2$. En effet, comme $d_G(u, y) = d_G(u, x) = k + 1$ et $d_G(u, y_1) = k + 2$, et comme $B_{k+1}(u)$ est convexe, $y \sim x$. Si x était distinct de y_2 , alors G contiendrait un W_4 induit formé par $\{y, y_2, x, z, y_1\}$ (centré en y_1). \square

Remarquons que le lemme 43 précédent ne traite pas le cas où $F(x)$ et $F(y)$ sont tous deux des cônes ou tous deux des panneaux. Ce cas est en fait impossible d'après le lemme qui suit.

Lemme 44. *Un sommet u dans un cône $F(x)$ ne peut pas être adjacent à un sommet v dans un cône $F(y) \neq F(x)$. De même, un sommet $v \neq y$ dans un panneau $F(y)$ ne peut pas être voisin d'un sommet u dans un panneau $F(x) \neq F(y)$.*

Démonstration. Nous démontrons les deux propriétés par l'absurde : $F(x)$ et $F(y)$ sont supposées être deux fibres de même nature. Remarquons tout d'abord que cette hypothèse implique que $d_G(u, x) = d_G(v, y) =: k$. En effet, $d_G(u, x) > k$ impliquerait que $y \in \text{Pr}(u, \text{St}(z))$, et $d_G(u, x) < k$ impliquerait que $x \in \text{Pr}(v, \text{St}(z))$.

Commençons par supposer que $F(x)$ et $F(y)$ sont deux cônes. Alors $F(y)$ ne peut pas être 1-voisin de $F(x)$ (autrement, G contiendrait un K_4 ou un cycle isométrique de longueur supérieure à 3). $F(x)$ et $F(y)$ ne peuvent pas être 2-voisins, autrement il existerait un sommet $w \in \text{St}(z)$ adjacent à x et y , et alors $w \notin B_k(\{u, v\})$ impliquerait que $x \sim y$. Donc $F(x)$ et $F(y)$ sont r -voisins pour un certain $r > 2$. Par la condition du triangle, il existe un sommet $t \sim u, v$ à distance $k + 1$ de z . Ce sommet t ne peut appartenir ni à $F(x)$, ni à $F(y)$. En

effet, t est à distance $k + 1$ de z et c'est un voisin de u et v . Si t appartenait à $F(x)$ (ou, de façon similaire, à $F(y)$), alors il serait à distance $k - 1$ de x , et donc v serait à distance k de x . Mais alors, soit x et y devraient coïncider (ce que nos hypothèses ne permettent pas), soit v devrait posséder plus de deux projections sur $N_G(z)$ (ce qui est impossible par le lemme 42). Posons $\text{Pr}(u, N(z)) := \{x_1, x_2\}$ et $\{y_1, y_2\} := \text{Pr}(v, N(z))$, et considérons un $w \in \text{Pr}(t, N(z))$ distinct de x_1 et de y_1 . Alors, puisque $d_G(u, w) = d_G(v, w) = d_G(u, x_1) = d_G(v, y_1) = k + 1$ et $z \notin B_{k+1}(u) \cup B_{k+1}(v)$, nous pouvons déduire que $w \sim x_1, y_1$. Comme u et v ne peuvent pas être à distance $k + 1$ de plus de deux sommets de $\text{St}(z)$ chacun, $x_2 = w = y_2$. $F(x)$ et $F(y)$ sont donc 2-voisins, et nous obtenons une contradiction.

Pour la second assertion, considérons que $F(x)$ et $F(y)$ dénotent deux panneaux. Comme $d_G(u, x) = d_G(v, y) = k$, nous pouvons remarquer que $x \sim y$ de par la convexité de la boule $B_k(\{u, v\})$, à laquelle z n'appartient pas. u et v étant adjacents et à distance $k + 1$ de z , il existe un sommet $t \sim u, v$ avec $d_G(t, z) = k$. Ce sommet t ne peut appartenir ni à $F(x)$, ni à $F(y)$ car, si t appartenait à $F(x)$ (resp. $F(y)$), alors v (resp. u) serait à distance k de x et de y (u ou v n'appartiendrait donc pas à un panneau, mais à un cône). Puisque $d_G(t, x) = d_G(t, y) = k - 1$, il existe $w \sim x, y$ à distance $k - 2$ de t . Donc t appartient à un cône $F(w)$ avec $w \sim x, y$. De l'égalité $d_G(t, z) = k$ découle que $d_G(t, w) = k - 2$. mais alors, $d_G(u, y) \leq 1 + d_G(t, w) + 1 = k$ conduit à $u \in F(w)$, ce qui est une contradiction. \square

En d'autres mots, le lemme utilitaire 44 établit que, si u et v sont deux sommets adjacents de $V \setminus \text{St}(z)$ appartenant à deux fibres distinctes, alors l'une de ces fibres doit être un cône et l'autre un panneau. Ainsi, lorsque, dans certaines preuves, nous nous intéresserons à des arêtes entre deux fibres, ce lemme 44 nous permettra d'éviter les cas (impossibles) dans lesquels un sommet dans un panneau possède un voisin dans un autre panneau, ou un sommet dans un cône possède un voisin dans un autre cône.

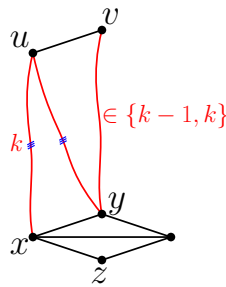


FIGURE 8.4 – Illustration du lemme 44.

Lemme 45. *Chaque fibre $F(x)$ de $\text{St}(z)$ est étoilée (par rapport à x).*

Démonstration. Soient $F(x)$ une fibre de $\text{St}(z)$, $u \in F(x)$ et $v \in I(u, x)$. Si $F(x)$ est un cône, alors $v \in I(u, x)$ et $x \in I(u, z)$ impliquent que $v \in I(u, z)$. Cela signifie

forcément que $v \in F(x)$. En effet, supposons que $v \in F(y) \neq F(x)$ et considérons deux sommets $u' \in F(x) \cap I(u, v)$ et $v' \in F(y) \cap I(u, v)$ avec $u' \sim v'$. Alors, par le lemme 43, $d_G(v', x) \in \{d_G(u', x), d_G(u', x) + 1\}$. Cela contredit $v' \in I(u, x)$. Si $F(x)$ est un panneau et si v appartient à une fibre $F(y) \neq F(x)$, alors $F(y)$ doit être un cône 1-voisin de $F(x)$ (lemme 44). Par conséquent, y doit appartenir à un plus court (u, x) -chemin passant par v . Pourtant, $d_G(u, y) = d_G(u, x)$. \square

Lemme 46. *Chaque fibre de $\text{St}(z)$ induit un sous-graphe isométrique de G .*

Démonstration. Soit $F(x)$ une fibre de $\text{St}(z)$ et soient u et v deux sommets de $F(x)$. Considérons un quasi-médian (u', v', x') de (u, v, x) . $F(x)$ étant étoilée (d'après le lemme 45), les intervalles $I(u', x')$, $I(x, x')$, $I(u, u')$ et $I(v, v')$ sont tous contenus dans cette fibre. Afin de montrer que $F(x)$ est isométrique, nous devons prouver que le chemin entre u' et v' dans le triangle métrique (u', v', x') (c.-à-d., l'enveloppe convexe de (u', v', w')) ne sort pas de $F(x)$ non plus. Nous faisons cette preuve par contradiction. Pour simplifier les notations, commençons par remarquer les deux choses suivantes. Premièrement, puisque $I(u, u'), I(v, v') \subseteq F(x)$, nous pouvons supposer que $u = u'$ et $v = v'$. Deuxièmement, nous pouvons supposer que (u, v, x') constitue un contre-exemple minimal. Nous définissons cette minimalité vis-à-vis de deux critères : *base* et *hauteur*. Soit w le sommet le plus proche de u sur le (u, v) chemin du triangle métrique tel que $w \notin F(x)$. La minimalité de la base signifie que l'on suppose u adjacent à w (sinon nous remplaçons u par le voisin de w dans $F(x)$). De par la structure des triangles métriques établie dans le lemme 38, nous savons qu'il existe un sommet $u_1 \sim w, u$ à distance $k - 1$ de x , et que $u_1 \in F(x)$. Considérons le voisin w_1 de u_1 dans (u, v, x') à distance $k - 1$ de x . Si $w_1 \notin F(x)$, alors nous remplaçons (u, v, x') par (u_1, v_1, x') , où $v_1 \sim v$ est le voisin de v dans (u, v, x') à distance $k - 1$ de x . Cela définit la minimalité de la hauteur. Donc $u_1, w_1 \in F(x)$, et si le chemin entre u_1 et v_1 sort de $F(x)$, alors nous pouvons aussi remplacer (u, v, x') par (u_1, v_1, x') et ré-appliquer la minimalité de la base. En ré-appliquant tant que c'est possible les critères de base et de hauteur, nous pouvons alors supposer que le triangle métrique (u_1, v_1, x') est entièrement contenu dans $F(x)$ (voir la figure 8.5). En particulier, $w_1 \in F(x)$. Deux cas sont à considérer.

Cas 1. $F(x)$ est un panneau. Alors w appartient à un cône 1-voisin de $F(y)$ par le lemme 44. Comme $d_G(w, x) = k$, $d_G(w, y) = k - 1$. Par le lemme 43, nous avons de plus que $d_G(w_1, y) \in \{k - 1, k - 2\}$, et puisque $d_G(w, y) = k - 1$, nous pouvons conclure que $d_G(w_1, y) = k - 1$. Par la condition du triangle (appliquée à w, w_1 et y), il existe un sommet $t \sim w, w_1$ à distance $k - 2$ de y . Comme $F(y)$ est étoilée, et comme $t \in I(w, y)$, $t \in F(y)$. Cela signifie aussi que $d_G(t, x) = k - 1$. Par convexité de $B_{k-1}(x)$, t doit donc coïncider avec u_1 ou w_1 (autrement, $\{u_1, w_1, w, t\}$ induirait un K_4). Par les hypothèses de minimalité, u_1 et w_1 doivent appartenir à $F(x)$, mais alors $t \in F(x)$. Nous obtenons donc une contradiction.

Cas 2. $F(x)$ est un cône. Alors w appartient à un panneau 1-voisin $F(y)$, et ce cas est assez similaire au précédent. Par le lemme 43, $d_G(w, y) = d_G(w, x) = k$. Nous avons donc $d_G(w, y) = d_G(w_1, y) = k$, et il existe un sommet $t \sim w, w_1$ à distance $k - 1$ de y . Toujours en nous appuyant sur le lemme 43, nous déduisons $d_G(t, x) = k - 1$. Étant donné que $F(y)$ est étoilée, nous obtenons que $t \in F(y)$, et par la convexité de $B_{k-1}(x)$, nous concluons que $t = w_1$ or $t = u_1$. Au final, $t \in F(x)$. \square

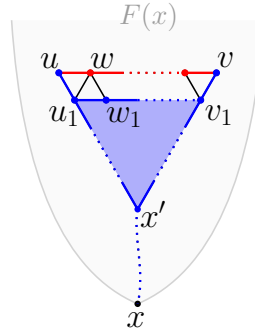


FIGURE 8.5 – Illustration de la preuve du lemme 46. Les hypothèses de minimalité impliquent que la partie bleue appartient à la fibre $F(x)$. Il s'agit alors de montrer que le plus court (u, v) -chemin rouge appartient aussi à $F(x)$.

Corollaire 10. *Chaque fibre de $\text{St}(z)$ induit un sous-graphe ponté de G .*

Démonstration. Cela découle directement du lemme 46 et du fait que les sous-graphes isométriques d'un graphe ponté soient eux-mêmes pontés. En effet, si H est un sous-graphe isométrique de G , alors chaque cycle isométrique de H est isométrique dans G . Puisque G ne contient pas de cycle isométrique de longueur supérieure à 3, H n'en contient pas non plus. \square

Lemme 47. *Soit $m \in V$ un sommet médian de G . Alors, pour tout $x \in \text{St}(m)$, $|F(x)| \leq |V|/2$.*

Démonstration. Supposons, par l'absurde, qu'il existe une fibre $F(x)$, $x \in \text{St}(m)$ telle que $|F(x)| \geq \frac{|V|+1}{2}$. Soit $m' \in I(m, x)$ un voisin de m (potentiellement x). Remarquons que les sommets de $F(x)$ appartiennent en particulier à $W(m', m)$. Par conséquent,

$$\begin{aligned} M(m) &= \sum_{v \in V} d_G(m, v) = M(m') + |W(m', m)| - |W(m, m')| \\ &\geq M(m') + \frac{|V| + 1}{2} - \frac{|V| - 1}{2} \end{aligned}$$

Mais alors $M(m') < M(m)$ contredit le choix de m comme sommet médian de G . \square

8.5.1 Frontières et bordures

Soit $z \in V$, et soient x et y deux sommets de $\text{St}(z)$. Les notions de frontières et bordures de fibres sont identiques à celles introduites pour les graphes médians dans le chapitre 7. C'est-à-dire que la frontière $\partial_{F(y)}F(x)$ de $F(x)$ vers $F(y)$ est l'ensemble $\{u \in F(x) : \exists v \in F(y) \text{ t.q. } uv \in E\}$ des sommets de $F(x)$ ayant un voisin dans $F(y)$. La bordure $\partial F(x)$ de $F(x)$ est l'union de ses frontières.

Lemme 48. *La bordure $\partial^*F(x)$ d'une fibre $F(x)$ de $\text{St}(z)$ est toujours un arbre étoilé.*

Démonstration. Nous distinguons deux cas, selon que $F(x)$ soit un cône ou un panneau.

Supposons, dans un premier temps, que $F(x)$ soit un cône. Alors x est à distance 2 de z , et x et z ont deux voisins communs y et y' . Pour prouver que $\partial^*F(x)$ est un arbre étoilé, il suffit de montrer que (1) pour tout sommet $v \in \partial^*F(x)$ l'intervalle $I(v, x)$ appartient à $\partial^*F(x)$, et (2), que v possède un unique voisin dans $I(v, x)$. Soit u un voisin de v dans une fibre 1-voisine de $F(x)$. Alors u appartient forcément à un panneau (par le lemme 44) : donc $F(y)$ ou $F(y')$. Disons $u \in F(y)$. Supposons aussi que u est le voisin de v le plus proche de y (dans $F(y)$). Soit $k := d_G(v, x)$. De par la définition d'un cône, $d_G(v, y) = d_G(v, y') = k + 1$ et $d_G(v, z) = k + 2$. Cela implique que $d_G(u, y) \geq k$. Comme u appartient au panneau $F(y)$, $d_G(u, x) \geq d_G(u, y)$. Soit u' un voisin quelconque de u dans $I(u, y)$. Si $d_G(u, y) = k + 1$, puisque $d_G(u', y) = d_G(v, x) = k$ et $d_G(u, x) \geq d_G(u, y) = k + 1$, nous concluons de la convexité de $B_k(\{x, y\})$ que $u' \sim v$, contrairement au choix fait de u . Donc $d_G(u, y) = k$. Choisissons n'importe quel voisin w de v dans $I(v, x)$. Alors $d_G(y, w) = d_G(y, u) = k$ et $d_G(y, v) = k + 1$. $B_k(y)$ étant convexe, nous déduisons que $w \sim u$, ce qui conduit à $w \in \partial^*F(x)$. Si v possède un second voisin w' dans $I(v, x)$ alors, par le même argument $u \sim w'$. De plus, $w \sim w'$ par la convexité de $B_{k-1}(x)$, car $d_G(x, w) = d_G(x, w') = k - 1$, $d_G(x, v) = k$ et $v \sim w, w'$. Par conséquent, les sommets u, v, w et w' induisent un K_4 . Ainsi, $I(v, x)$ consiste en un simple chemin, inclus dans $\partial^*F(x)$.

Notons aussi que l'unique voisin w de v dans $I(v, x)$ doit être adjacent à tout voisin u' de u dans $I(u, y)$. En effet, $u', w \in B_{k-1}(\{x, y\})$. Et comme $u \sim u', w$ et $d_G(u, x) = d_G(u, y) = k$, par la convexité de $B_{k-1}(\{x, y\})$ nous déduisons que $u' \sim w$.

Supposons maintenant que $F(x)$ soit un panneau de $\text{St}(z)$. Choisissons un sommet quelconque $u \in \partial^*F(x)$. Nous devons montrer que $I(u, x) \subseteq \partial^*F(x)$ et

que u possède un unique voisin dans $I(u, x)$. Soit v un voisin de u dans une fibre $F(y)$ 1-voisine de $F(x)$. $F(y)$ est donc un cône tel que $y \sim x, x'$, avec $x' \sim x$ et $z \sim x, x'$. Supposons que v soit le voisin de u dans $F(y)$ le plus proche de y . Soit $d_G(v, y) := k$. Par le lemme 43, nous savons que $d_G(u, x) \in \{k, k + 1\}$. Si $d_G(u, x) = k$, alors nous déduisons que u est adjacent au voisin w de v dans $I(v, y)$, contrairement à l'hypothèse de minimalité de v . Donc $d_G(u, x) = k + 1$. Dans ce cas, si u' est un voisin de u dans $I(u, x)$, alors $d_G(u', x) = d_G(v, y) = k$ et $d_G(u, y) \geq d_G(u, x) = k + 1$. Par la convexité de $B_k(\{y, x\})$, nous concluons que $u' \sim v$. Cela implique que, si u possède deux voisins u' et u'' dans $I(u, x)$, alors u, u', u'' et v induisent un K_4 (interdit), car alors $u' \sim u''$ par la convexité de $B_k(x)$ qui contient u' et u'' , mais pas u . Par conséquent, $I(u, x)$ est un simple chemin inclus dans $\partial^*F(x)$. \square

Remarque 29. La bordure d'une fibre est un arbre étoilé, mais ce n'est pas toujours un arbre induit.

Le résultat suivant peut alors être obtenu en tant que corollaire du lemme 48.

Corollaire 11. *Soit x un sommet quelconque de $\text{St}(z)$. Pour toute paire de sommets u et v de $\partial^*F(x)$, $d_G(u, v) \leq d_{\partial^*F(x)}(u, v) \leq 2 \cdot d_G(u, v)$.*

Démonstration. Pour un $x \in \text{St}(z)$, considérons deux sommets u et v de $T := \partial^*F(x)$ ainsi qu'un quasi-médian (u', v', x') de (u, v, x) . Alors $d_G(u, v) = d_G(u, u') + d_G(u', v') + d_G(v', v)$. Puisque T est un arbre étoilé par le lemme 48, $d_T(u, u') = d_G(u, u')$ et $d_T(v, v') = d_G(v, v')$, et $d_T(u', x') = d_G(u', x')$ et $d_T(x', v') = d_G(x', v')$. Remarquons de plus que $x' \in I(u, x) \cap I(v, x)$ implique que x' soit le plus proche ancêtre commun de u et v dans T . Comme les triangles métriques sont équilatéraux dans les graphes pontés, $d_T(u', x') + d_T(x', v') = 2d_G(u', v')$. \square

Remarque 30. Si G est δ -hyperbolique, alors l'erreur multiplicative de 2 sur les distances dans le corollaire 11 devient une erreur additive de 2δ .

8.5.2 Projections et représentants

Nous décrivons ici la structure des projections des sommets sur les bordures de leurs fibres voisines. Nous justifions ensuite que les sommets dans les panneaux n'auront qu'un nombre constant de "représentants" sur leur bordure, bien que ces panneaux soient 1-voisins d'un nombre arbitraire de cônes. C'est l'objet du lemme 51, similaire dans le principe au lemme 32 pour les graphes médians sans cube.

Lemme 49. *Soient $F(x)$ une fibre et $u \in V \setminus F(x)$. Alors la projection métrique $\Pi := Pr(u, F(x)) = Pr(u, \partial^*F(x))$ est un arbre induit de G .*

Démonstration. Puisque la projection métrique de u sur $F(x)$ appartient à une frontière qui est elle-même un arbre étoilé (par le lemme 48, Π est une forêt étoilée (c.-à-d., un ensemble d'arbres étoilés). Nous allons donc montrer qu'elle est connexe dans $T := \partial^* F(x)$ (c.-à-d., Π est un arbre étoilé). Pour ce faire, nous devons en fait montrer la propriété plus forte que deux sommets a et b dans des branches distinctes de Π ne peuvent pas être reliés par une arête de G . (c.-à-d., Π est un arbre induit de G). Soient $v, w \in \Pi$. Nous supposons par l'absurde que v et w ne soient pas connectés par un chemin dans T . Puisque T est un arbre étoilé (et enraciné en x), deux cas sont à considérer selon les positions respectives de v et w dans T .

Supposons, tout d'abord que w soit un ancêtre de v (ou vice-versa). Alors tous les sommets de la branche de v entre v et w sont sur un plus court (v, w) -chemin (car T est étoilé) et sont à distance au moins $k + 1 := d_G(u, v) + 1 = d_G(u, w) + 1$ de u (puisque'ils ne sont pas dans sa projection), cela contredit la convexité de la boule $B_k(u)$. Rappelons que chaque branche de T est induite dans G , nous vérifions donc bien, jusqu'ici que Π est un arbre induit dans G .

Considérons maintenant que v et w soient sur des branches distinctes de T , et appelons t leur plus proche ancêtre commun. Soit (v', w', t') un quasi-médian de v, w et t . Du fait que T soit un arbre étoilé, que $t' \in I(v, t) \cap I(w, t)$ et que t soit le plus proche ancêtre commun de v et w , nous déduisons que $t = t'$. Nous affirmons que tous les sommets de T entre v et v' , entre v' et w' , et entre w' et w appartiennent à Π . En effet, ils sont tous sur un plus court (v, w) -chemin. Or, la boule $B_k(u)$ est convexe et $v, w \in B_k(u)$. Comme tous ces sommets appartiennent aussi à T , nous en déduisons qu'ils sont tous à distance k de u . Nous voulons maintenant montrer que $v' = w' = t$, cela conduira à une contradiction avec le fait que v et w ne soient pas connectés dans T . Supposons, par l'absurde, que $v' \neq w'$ en considérant une arête ab sur le chemin entre v' et w' dans (v', w', t) . Rappelons que, d'après le théorème 10, $d_G(a, x) = d_G(b, x) =: l$. Puisque $a \sim b$ et $d_G(a, u) = d_G(b, u) = k$, il doit exister un sommet $c \sim a, b$ à distance $k - 1$ de u , et ce sommet est dans une fibre $F(y)$ voisine de $F(x)$. Par le lemme 44, l'une des fibres doit être un cône et l'autre un panneau.

Obtenons d'abord une contradiction dans le cas $F(x)$ panneau et $F(y)$ cône. Par la condition du triangle appliquée à a, b et x , il existe un sommet $d \sim a, b$ à distance $l - 1$ de x . Par le lemme 43, deux sous-cas sont à considérer : $d_G(c, y) = l$ et $d_G(c, y) = l - 1$. Si $d_G(c, y) = l$, alors $d_G(c, y) = d_G(y, d) = l$ et $d_G(b, y) = l + 1$. La convexité de $B_l(y)$ implique donc $c \sim d$, et alors $\{a, b, c, d\}$ induit un K_4 . Si $d_G(c, y) = l - 1$, considérons le sommet $x' \sim y, x$ dans $I(z, y)$. Alors $d_G(c, x') = d_G(d, x') = l$, mais $d_G(b, x') = l + 1$. De par la convexité de $B_l(x')$, nous déduisons que $c \sim d$. Donc $\{a, b, c, d\}$ induit un K_4 .

Supposons maintenant que $F(x)$ soit un cône et $F(y)$ un panneau. De nouveau

par le lemme 43, il nous faut considérer les sous-cas $d_G(c, y) = l$ et $d_G(c, y) = l+1$. Si $d_G(c, y) = l$, alors $d_G(c, y) = d_G(d, y) = l$ et $d_G(b, y) = l+1$ impliquent que $c \sim d$ de par la convexité de $B_l(y)$. Donc $\{a, b, c, d\}$ induit un K_4 . Enfin, si $d_G(c, y) = l+1$, considérons un sommet $e \sim c$ dans $F(y)$ sur un plus court (c, y) -chemin, et considérons le sommet $y' \sim x, y$ dans $I(x, z)$. Alors $d_G(b, y') = d_G(e, y') = l+1$ et $d_G(c, y') = l+2$. De la convexité de $B_{l+1}(y')$ suit que $e \sim b$. De la même façon, nous montrons que $a \sim e$. Par conséquent, $\{a, b, c, e\}$ induit un K_4 .

Nous obtenons donc, dans chaque cas (illustrés dans la figure 8.6), que l'existence d'une arête ab entre v' et w' est contradictoire. Donc $v' = w' = t$ et $t \in \Pi$. \square

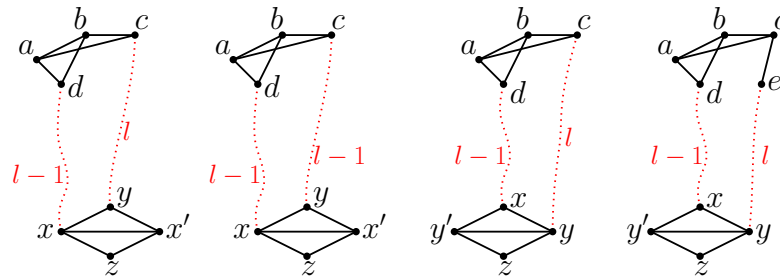


FIGURE 8.6 – Aide à la lecture des quatre cas pour montrer $t \in \Pi$ dans la preuve du lemme 49.

Lemme 50. Soient $F(x)$ une fibre et $u \in V \setminus F(x)$. Il existe un unique sommet u' dans $\Pi := Pr(u, F(x))$ à distance minimum de x . De plus, $d_\Pi(u', v) \leq d_G(u, u') = d_G(u, v)$ pour tout $v \in \Pi$ (où $d_\Pi(u', v)$ est un abus de notation pour $d_{G(\Pi)}(u', v)$).

Démonstration. L'unicité de u' est due au fait que Π soit un sous-arbre enraciné de l'arbre étoilé (par rapport à x) $T := \partial^* F(x)$. En effet, toute paire de sommets (a, b) de Π possède un plus proche ancêtre commun dans Π qui coïncide avec le plus proche ancêtre commun dans T , or ce dernier doit être plus proche de x que a et b (ou à égale distance si $a = x$ ou $b = x$).

Considérons un sommet $v \in \Pi$. L'égalité $k := d_G(u, u') = d_G(u, v)$ est vérifiée par définition de Π en tant que projection métrique u sur $F(x)$. Supposons, par l'absurde, que $d_G(u', v) \geq k+1$. Alors $I(v, u') =: P$ est un chemin montant de longueur au moins $k+1$ dans un arbre étoilé. Par la condition des triangles appliquée à u et à chaque paire de sommets voisins de P , nous dérivons au moins k sommets. Par des arguments similaires à ceux fournis dans la preuve du lemme 38, nous pouvons montrer que chacun de ces (au moins k) sommets est distinct des autres (sinon, nous aurions un raccourci dans P). Toujours comme dans la preuve du lemme 38, nous pouvons montrer que ces sommets forment un chemin et, par induction sur ce nouveau chemin plus court, nous pouvons montrer qu'alors (u, v, t) forme un triangle métrique non-équilatéral, ce qui n'est pas possible. \square

Autrement formulé, le lemme 50 établit que chaque branche de l'arbre Π est de longueur inférieure ou égale à $d_G(u, u')$.

Lemme 51. *Soit u un sommet d'un graphe ponté $G = (V, E)$, sans K_4 , et soit T un arbre étoilé enraciné en $r \in V$. Soient u_1 et u_2 les deux sommets extrémaux par rapport à r des deux chemins montants de $I(u, r) \cap T$ (il y en a au plus deux par le lemme 41). Alors, pour tout $v \in T$,*

$$\min\{d_G(u, u_1) + d_T(u_1, v), d_G(u, u_2) + d_T(u_2, v)\} \leq 2 \cdot d_G(u, v)$$

Démonstration. Supposons que $\min_{i \in \{1, 2\}} \{d_G(u, u_i) + d_T(u_i, v)\}$ soit atteint pour $i = 1$. Soit $x \in T$ le plus proche ancêtre commun de u_1 et v . Notons que, par le lemme 41, nous savons que $v \notin I(u, r)$, à moins que $v = x$. Mais nous pouvons supposer que $v \neq x$, car autrement nous aurions directement $d_G(u, v) = d_G(u, u_1) + d_T(u_1, v)$. Considérons un quasi-médian (u', v', r') du triplet (u, v, r) (voir figure 8.7). Nous pouvons faire les deux remarques suivantes :

(1) T étant étoilé, $I(r, v)$ est l'une de ses branches, et v' appartient donc à cette branche. Si $v' \in I(r, x)$, alors $v' \in I(u_1, r) \cap I(v, r)$ implique que $v' = x$ et que $d_G(u, u_1) + d_T(u_1, v') + d_T(v', v) = d_G(u, v)$. En effet, si $v' = x$, alors $v' \in I(u, r)$ (car $x \in I(u, r)$). Donc $d_G(u, u_1) + d_G(u_1, v') = d_G(u, v')$. Aussi, comme T est étoilé, $d_G(u_1, v') = d_T(u_1, v')$ et $d_T(v', v) = d_G(v', v)$. Enfin, comme v' est dans le quasi-médian entre u et v , il est sur un plus court (u, v) -chemin. Donc $d_G(u, u_1) + d_T(u_1, v') + d_T(v', v) = d_G(u, v') + d_G(v', v) = d_G(u, v)$. Considérons donc que $v' \in I(x, v)$.

(2) T étoilé et $r' \in I(v, r)$ impliquent que r' est sur la branche $I(v, r)$ de T . Mais comme $r' \in I(u, r)$, et comme $I(x, v) \cap I(u, r) = \{x\}$, nous en déduisons que r' est entre r et x . Comme $r' \in I(u', r) \cap I(u_1, r)$, et $u_1, u' \in I(u, r)$, nous montrons que $d_G(u, u') + d_G(u', r') = d_G(u, u_1) + d_T(u_1, r')$. En effet, $u' \in I(u, r')$ car il est dans le quasi-médian de (u, v, r) ; $u_1 \in I(u, r')$ par définition, donc les distances entre u et r' en passant par u_1 et en passant par u' sont égales : $d_G(u, u') + d_G(u', r') = d_G(u, u_1) + d_G(u_1, r')$. Enfin, comme T est étoilé (et r' appartient à T), $d_G(u_1, r') = d_T(u_1, r')$.

Du fait que le quasi-médian (u', v', r') soit équilatéral, nous avons aussi

$$\begin{aligned} 2 \cdot d_G(u, v) &\geq d_G(u, u') + d_G(u', r') + d_T(r', v) \\ &= d_G(u, u_1) + d_G(u_1, r') + d_T(r', v) \\ &= d_G(u, u_1) + d_T(u_1, v) \end{aligned}$$

□

En termes plus informels, si u appartient à une fibre $F(x)$ et $T := \partial^* F(x)$, et si u stocke seulement des informations relatives à u_1 et u_2 , alors les plus courts

chemins vers chacun des autres sommets de T sont toujours “assez proche” d’un chemin passant par u_1 ou u_2 (des distances approchées entre u et les sommets de T peuvent donc être calculées avec des informations relatives seulement à ces deux représentants).

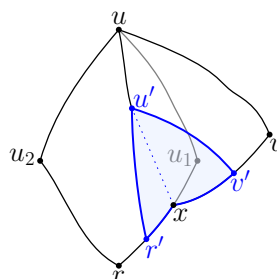


FIGURE 8.7 – Notations du lemme 51.

8.6 Classification des sommets et plus courts chemins

Nous présentons dans cette section les cas dans lesquels un plus court chemin entre deux sommets passe par le centre z de l'étoile $\text{St}(z)$ (lemme 52), ceux dans lesquels passer par ce sommet z peut conduire à une erreur multiplicative d'au plus 2 (lemme 53), et dans quels cas nous pourrions faire une erreur multiplicative d'au plus 4 (lemme 54).

Lemme 52. Soient $F(x)$ et $F(y)$ deux fibres, et soient $u \in F(x)$ et $v \in F(y)$. Alors $z \in I(u, v)$ si et seulement si $F(x)$ et $F(y)$ sont distinctes et (au choix) :

- (i) deux panneaux k -voisins, pour $k \geq 2$;
- (ii) un panneau et un cône k -voisins, pour $k \geq 3$;
- (iii) deux cônes k -voisins, pour $k \geq 4$;

Démonstration. Considérons un quasi-médian (u', v', z') de (u, v, z) . z appartient à un plus court (u, v) -chemin si et seulement si $u' = v' = z' = z$. Dans ce cas, soient $s \in I(u, z)$ et $t \in I(v, z)$ deux voisins de z . Puisque z appartient à un plus court (u, v) -chemin, s et t ne peuvent pas être adjacents. Il s'en suit (voir la figure 8.8) que $F(x)$ et $F(y)$ sont k -voisins avec :

- (1) $k \geq 2$ si $F(x)$ et $F(y)$ sont tous deux des panneaux ;
- (2) $k \geq 3$ si l'un de $F(x)$ et $F(y)$ est un cône et l'autre est un panneau ;
- (3) $k \geq 4$ si $F(x)$ et $F(y)$ sont tous deux deux cônes.

Pour l'implication inverse, considérons les cas dans lesquels z n'appartient pas à un plus court (u, v) -chemin. Commençons par remarquer que, si $F(x) = F(y)$, alors z ne peut pas appartenir à un tel plus court chemin, car alors $d_G(u, v) \leq d_G(u, x) + d_G(x, v) < d_G(u, z) + d_G(z, v)$. Nous supposons donc à présent que $F(x) \neq F(y)$. Trois cas sont à observer selon les statuts de $F(x)$ de $F(y)$.

Cas 1. $F(x)$ et $F(y)$ sont tous deux des panneaux. Si $z = z'$ alors, d'après le lemme 39, x et y doivent être les deux voisins de z respectivement sur les plus courts (z, u') -chemin et (z, v') -chemin, et $x \sim y$ (c.-à-d., $F(x)$ et $F(y)$ sont 1-voisins). Si $z \neq z'$, considérons un sommet $z'' \in I(z, z')$ adjacent à z . Alors $z'', x \in I(u, z)$ et, comme u appartient à un panneau, $z'' = x$. Avec le même argument, nous obtenons, $z'' = y$. Par conséquent $F(x) = F(y)$ alors que nous les avons supposés distincts.

Cas 2. $F(x)$ est un cône et $F(y)$ est un panneau (le cas symétrique étant similaire). Dénotons par x' et x'' les deux voisins de x dans l'intervalle $I(x, z)$. Si $z = z'$, encore par le lemme 39, y et x' (ou x'') doivent appartenir au quasi-médian, et alors $x' \sim y$ (ou $x'' \sim y$). Par conséquent, $F(x)$ et $F(y)$ sont 2-voisins. Si $z \neq z'$, considérons à nouveau un voisin z'' de z dans $I(z, z')$. Puisque $x', x'' \in I(u, z)$, z'' doit coïncider avec x' ou x'' , disons x' . De même, $z'' = y$. Ainsi $F(x)$ et $F(y)$ sont 1-voisins.

Cas 3. $F(x)$ et $F(y)$ sont tous deux des cônes. Dénotons par x' et x'' les deux voisins de x dans $I(x, z)$, et par y' et y'' ceux de y . Encore une fois, si $z = z'$, alors x' (ou x'') et y' (ou y'') appartiennent au quasi-médian (u', v', z') , cela conduisant à $x' \sim y'$ et donc au fait que $F(x)$ et $F(y)$ soient 3-voisins. Si $z \neq z'$, considérons $z'' \in I(z, z')$, $z'' \sim z$. Par les mêmes arguments que dans les cas précédents, nous obtenons que $z'' = x' = y'$ (à un renommage près des sommets x'' et y''). Donc $F(x)$ et $F(y)$ sont 2-voisins. \square

Nous donnons maintenant une classification des paires de sommets de manière similaire à ce que nous avons fait pour les graphes médians sans cubes du chapitre précédent. Soient $z \in V$, $x, y \in \text{St}(z)$ et $(u, v) \in F(x) \times F(y)$. Alors u et v sont

- *proches* si $F(x) = F(y)$;
- *séparés* si $z \in I(u, v)$;
- *1pp-voisins* si $F(x)$ et $F(y)$ sont panneaux 1-voisins ;
- *1pc-voisins* si $F(x)$ et $F(y)$ sont 1-voisines, l'une d'elles étant un panneau et l'autre un cône ;
- *2cc-voisins* si $F(x)$ et $F(y)$ sont deux cônes 2-voisins ;
- *2pc-voisins* si $F(x)$ et $F(y)$ sont 2-voisines, l'une d'elles étant un panneau et l'autre un cône ;
- *3cc-voisins* si $F(x)$ et $F(y)$ sont deux cônes 3-voisins.

Lemme 53. *Soient u et v deux sommets ni séparés, ni 1pc-voisins. Alors*

$$d_G(u, v) \leq d_G(u, z) + d_G(z, v) \leq 2 \cdot d(u, v).$$

Démonstration. Notons $F(x)$ et $F(y)$ les fibres contenant respectivement u et v . Considérons un quasi-médian (u', v', z') de u, v et z . Nous voulons montrer que $z = z'$.

Supposons dans un premier temps que u et v soient 1pp-voisins. Si x et y sont sur un plus court chemin, alors $z = z'$. Supposons donc que ce ne soit pas le cas. Alors il existe un cône $F(w) \sim F(x), F(y)$ tel que $I(u, v) \cap F(w) \neq \emptyset$. Nous affirmons que, si $z' \notin F(w)$, alors $z = z'$. En effet, cela découle directement du fait que $z' \in I(u, z) \cap I(v, z)$, $x \notin I(v, z)$ et $y \notin I(u, z)$. Montrons donc que $z' \notin F(w)$. Remarquons pour cela que $x \notin I(v, z)$, $y \notin I(u, z)$ et $x, y \in I(w, z)$ impliquent que $w \notin I(u, z) \cup I(v, z)$. Par définition de (u', v', z') , nous avons, en particulier, $z' \in I(u, z)$. De plus, $z' \in F(w)$ impliquerait la contradiction $w \in I(z', z) \subseteq I(u, z)$.

Considérons maintenant que u et v sont 2pc-voisins. Disons que $F(y)$ est un panneau, et nommons x_1 et x_2 les deux voisins de x dans $I(x, z)$. Nous montrons de la même manière que $z = z'$. En effet, $z' \in I(u, z)$ nécessite que $z' \in F(x) \cup F(x_1) \cup F(x_2) \cup \{z\}$, et $z' \in I(v, z)$ nécessite que $z' \in F(y) \cup \{z\}$. Nous en concluons que $z = z'$.

Les mêmes arguments permettent d'établir que, si u et v sont 2cc-voisins ou 3cc-voisins, alors $z = z'$. □

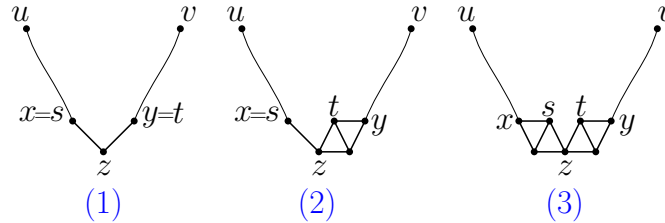


FIGURE 8.8 – Pour la preuve du lemme 52.

Remarque 31. Il est facile de remarquer que, si $F(x) = F(y)$, alors z n'appartient pas à un plus court (u, v) -chemin car $d_G(u, v) \leq d_G(u, x) + d_G(x, v) < d_G(u, z) + d_G(z, v)$.

Soient $F(x)$ un panneau et $F(y)$ un cône 1-voisin de $F(x)$. Nous posons $T := \partial^* F(x)$. Soient $u \in F(x)$ et $v \in F(y)$. Rappelons que $\Pi := \text{Pr}(v, T)$ induit un arbre d'après le lemme 49. Le sommet v' le plus proche de x dans l'arbre induit par Π est appelé un *représentant de v* sur la bordure T . De même, les deux sommets u_1 et u_2 décrits dans le lemme 51 sont appelés les *quasi-représentants de u* sur la

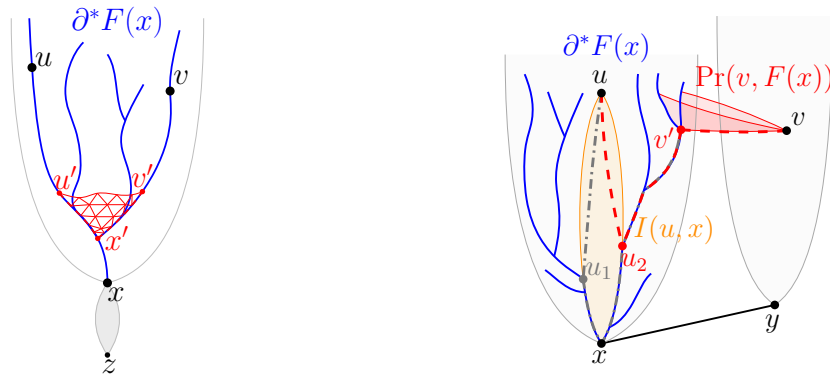


FIGURE 8.9 – À gauche, une illustration du corollaire 11. À droite, un dessin pour le lemme 54.

bordure T . Pour de tels u, v, u_1, u_2 et v' , le résultat suivant est vérifié (voir la figure 8.9, droite) :

Lemme 54.

$$\begin{aligned} d_G(u, v) &\leq \min\{d_G(u, u_1) + d_T(u_1, v'), d_G(u, u_2) + d_T(u_2, v')\} + d_G(v', v) \\ &\leq 4 \cdot d_G(u, v). \end{aligned}$$

Démonstration. Supposons que la valeur minimale $\min\{d_G(u, u_1) + d_T(u_1, v'), d_G(u, u_2) + d_T(u_2, v')\}$ soit atteinte pour u_1 . Soit $u' \in I(u, v) \cap T$ le plus proche possible de Π . Alors la distance exacte entre u et v est $d_G(u, u') + d_G(u', v)$, et nous devons comparer celle-ci à $d_G(u, u_1) + d_T(u_1, v') + d_G(v', v)$. L'inégalité triangulaire nous donne :

$$d_G(u, u_1) + d_T(u_1, v') \leq d_G(u, u_1) + d_T(u_1, u') + d_T(u', v').$$

Du fait que $v' \in \Pi$, nous obtenons :

$$d_G(v', v) \leq d_G(u', v).$$

Par le choix de u_1 et le lemme 51, nous savons aussi que :

$$d_G(u, u_1) + d_T(u_1, u') \leq 2 \cdot d_G(u, u').$$

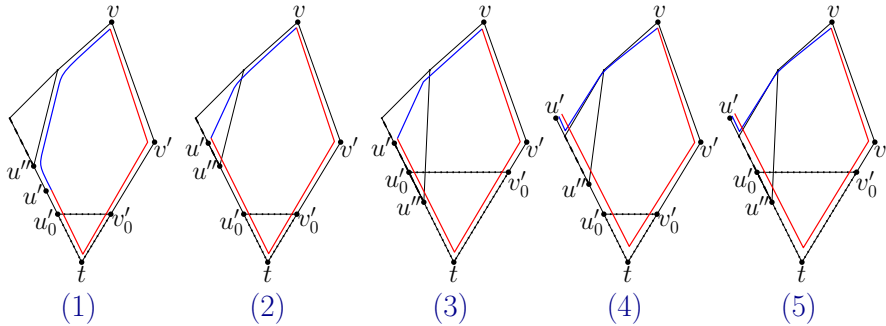
Il nous reste donc à comparer $d_T(u', v')$ avec $d_G(u', v)$.

Supposons pour commencer que u' et v' soient sur une même branche de T , et considérons un quasi-médian (u'_0, v'_0, v_0) de (u', v', v) . Alors $I(u', v') \subseteq T$ (car T est étoilé) implique que $u'_0, v'_0 \in T$. u' étant le plus proche possible de Π , nous avons $u' = u'_0$. Et comme $v'_0 \in I(v, v')$, $v'_0 \in \Pi$. Puisque v' est le plus proche possible de u' , $v' = v'_0$. Enfin, puisque le quasi-médian est équilatéral, nous déduisons que

$d_G(u, u') = d_G(u, v')$. Donc $u' \in \Pi$. Nous en déduisons que $d_T(u', v') \leq d_G(u', v)$ (lemme 50), ainsi que l'encadrement

$$\begin{aligned} d_G(u, v) &\leq d_G(u, u_1) + d_T(u_1, v') + d_G(v', v) \\ &\leq d_G(u, u_1) + d_T(u_1, u') + d_T(u', v') + d_G(v', v) \\ &\leq 2 \cdot d_G(u, u') + d_G(u', v) + d_G(u', v) \\ &= 2 \cdot d_G(u, v). \end{aligned}$$

Supposons maintenant que u' et v' soient sur des branches distinctes de T , et dénotons par t leur plus proche ancêtre commun dans cet arbre étoilé. Soit u'' le sommet le plus proche de t de la projection métrique de v sur la branche de u' . Considérons un quasi-médian (u'_0, v'_0, t_0) de (u', v', t) . Remarquons que T étoilé, $t_0 \in I(u', t) \cap I(v', t)$, et t plus proche ancêtre commun de u' et v' impliquent que $t = t_0$. De plus, u'_0 et v'_0 sont respectivement sur les branches de u' et de v' . Nous distinguons alors les cinq cas décrits dans la figure suivante. Les traits bleus correspondent à la distance exacte $d_G(u', v)$, les rouges représentent la distance approchée que nous y comparons.



En fait, nous pouvons remarquer que notre erreur sera maximale si u' est l'extrémité opposée à u'' dans la projection de v sur la branche de u' . En effet, la variation de l'erreur que nous faisons dans chaque cas a lieu sur la portion de chemin entre u' et u'' qui n'est pas sur le plus court (u, v) -chemin. La longueur de cette portion est maximale dans ce cas. Nous voulons maintenant montrer les deux inégalités suivantes :

Affirmation 54.1. $d_G(t, v') \leq d_G(v, u'')$ et $d_G(t, u'') \leq d_G(v, v')$.

Preuve de l'affirmation. Notons, pour commencer, que u'' et v' appartiennent à $I(v, t)$. Notons aussi que $I(t, u'') \cap I(t, v') = \{t\}$, mais que les intervalles $I(v, u'')$ et $I(v, v')$ peuvent s'intersecter sur une portion de $I(v, s)$ (plus grande que $\{v\}$). Soit s le sommet le plus éloigné de v dans cette intersection (v et s peuvent éventuellement coïncider). Nous allons montrer que $d_G(s, u'') = d_G(v', t)$ et $d_G(s, v') = d_G(u'', t)$.

Considérons un sous-graphe D de $G(I(s, t))$ consistant en le cycle ∂D formé par un plus court (u'', t) -, (t, v') -, (v', s) - et (s, u'') -chemin, ainsi qu'en l'intérieur de ce cycle (voir l'illustration 8.10). Un tel sous-graphe est planaire (dans notre cas) et peut être appelé un *disque*. Nous supposons D minimal au sens suivant : D contient v, v', t et u'' , ainsi qu'un nombre minimal de faces. Nous pouvons remarquer que, dans notre cas, le voisinage fermé (dans D) des coins induit soit un K_3 (c.-à-d., un triangle), nous parlerons alors de *coins de type 1*, soit un K_4^- (c.-à-d., deux triangles collés par une arête), les *coins de type 2* (rappelons que la définition des coins est donnée dans la section 4.4). Notons tout d'abord que s et t sont des coins de type 1 puisqu'ils ont exactement deux voisins dans l'intervalle $I(s, t)$ par le lemme 40. Montrons maintenant que les seuls autres sommets pouvant être des coins sont u'' et v' . En effet, les chemins entre u'' et t et entre t et v' étant convexes (car ils sont dans des branches d'un arbre étoilé), ils ne peuvent pas contenir de coins (autres que u'', t ou v'). Pour ce qui est des (u'', s) - et (s, v') -chemins, supposons par l'absurde que l'un d'eux, disons celui entre u'' et s , contienne un coin (différent de u'' ou s). Si ce coin est de type 1, nous obtenons directement une contradiction avec le fait qu'il soit dans $I(u'', v)$. S'il est de type 2, dénotons-le par a et appelons a' sont voisin dans l'intérieur de D . Alors a et a' sont tout deux sur un plus court (u'', s) -chemin, or remplacer a par a' définit un disque avec deux triangles de moins, contredisant la minimalité de D . Donc u'' et v' sont les seuls coins supplémentaires éventuels. D'après le lemme 40 de structure de $I(s, t)$ et la minimalité de D , pour tout $w \in D \setminus \partial D$, $\kappa(w) = 0$ et, pour tout $w \in \partial D \setminus \{s, t, v', u''\}$, $\tau(w) = 0$. Puisque $\tau(s) = \tau(t) = \frac{2\pi}{3}$, d'après le théorème de Gauss-Bonnet (16), $\tau(u'') + \tau(v') = \frac{2\pi}{3}$. Donc soit v' et u'' sont tous deux de type 2, soit l'un d'eux est de type 1 et l'autre n'est pas un coin. Il est facile de constater que le second cas est impossible, puisque si l'un de ces sommets est de type 1, il ne peut pas être sur un plus court (s, t) -chemin. Donc u'' et v' sont de type 2. De par la structure de l'intervalle $I(s, t)$ décrite dans le lemme 40, nous concluons que $d_G(s, u'') = d_G(v', t)$ et que $d_G(s, v') = d_G(u'', t)$. ■

De $d_G(v, t) = d_G(v, u'') + d_G(u'', t) = d_G(v, v') + d_G(v', t) = d_G(v, u') + d_G(u'', t)$, nous déduisons que

$$d_G(v, v') + d_T(v', u') < 4 d_G(v, u')$$

Au final,

$$\begin{aligned}
 d_G(u, v) &\leq d_G(u, u_1) + d_T(u_1, u') + d_T(u', v') + d_G(v', v) \\
 &\leq 2 \cdot d_G(u, u') + 3 \cdot d_G(u', v) + d_G(u', v) \\
 &\leq 4 \cdot d_G(u, v).
 \end{aligned}$$

□

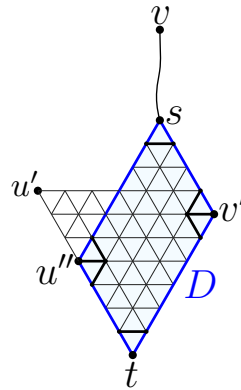


FIGURE 8.10 – Pour la preuve de l’affirmation 54.1.

8.7 Schéma de distance 4-approché pour les graphes pontés sans K_4

Nous décrivons maintenant un schéma d’étiquetage de distance 4-approché pour les graphes pontés sans K_4 , basé sur les propriétés établies dans les sections précédentes.

8.7.1 Encodage

Nous commençons cette section par une brève description de l’encodage que nous faisons des étoiles $\text{St}(m)$ d’un graphe ponté $G = (V, E)$ sans K_4 . Nous expliquerons dans la section 8.7.2 quel usage en fera la fonction de décodage des distances. Dans la section 8.7.1.2, nous décrivons les étiquettes $\text{LDA}(u)$ de distances 4-approchées des sommets $u \in V$ construites par l’algorithme d’encodage `ENC_DIST`. Nous n’y précisons pas les étapes de la procédure récursive, ces dernières étant très similaires à celles décrites dans le chapitre 7.

8.7.1.1 Encodage de l’étoile

Considérons un sommet m de G et l’étoile $\text{St}(m)$ de ce sommet. Les étiquettes des sommets u de $\text{St}(m)$ dans cette étoile seront notées $\text{L}_{\text{St}(m)}(u)$. Nous définissons $\text{L}_{\text{St}(m)}(m) := 0$ (où 0 sera interprété comme l’ensemble vide \emptyset). Les voisins de m prennent chacun une étiquette distincte dans $\{1, \dots, \text{deg}(m)\}$ (interprétée comme un singleton). L’étiquette $\text{L}_{\text{St}(m)}(u)$ d’un sommet u à distance 2 de m vaut

la concaténation des étiquettes $L_{\text{St}(m)}(u')$ et $L_{\text{St}(m)}(u'')$ des deux voisins u' et u'' de u dans $I(u, m)$ (c.-à-d., un 2-ensemble).

Remarque 32. Les sommets à distance 2 de m n'ont pas nécessairement un identifiant unique. De plus, l'étiquetage des sommets de l'étoile ne permet pas non plus de déterminer l'adjacence pour toute paire de sommets. En effet, rien ne permet de déterminer l'adjacence de deux sommets encodés par des singletons (le fait qu'ils soient des singletons indique seulement qu'ils sont voisin de m , voir la figure 8.11).

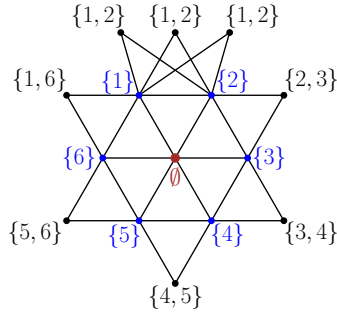


FIGURE 8.11 – Exemple de l'encodage des sommets d'une étoile $\text{St}(m)$ dans un graphe ponté sans K_4 .

8.7.1.2 Encodage d'un graphe ponté sans K_4

Soit u un sommet quelconque de G . Nous décrivons ici la partie d'étiquette $\text{LDA}_i(u)$ de u construite, à l'étape (réursive) i , par la procédure [ENC_DIST](#). Comme précédemment, $\text{LDA}_i(u)$ est constituée de trois parties : “St”, “1er” et “2nd”. La première partie, LDA_i^{St} , contient les informations relatives à l'étoile $\text{St}(m)$ du médian m de l'étape correspondante :

- (1) $\text{LDA}_i^{\text{St}[\text{Med}]}(u) := \text{id}(m)$;
- (2) $\text{LDA}_i^{\text{St}[\text{Dist}]}(u) := d_G(u, m)$;
- (3) $\text{LDA}_i^{\text{St}[\text{Porte}]}(u) := L_{\text{St}(x)}(m)$, où $x \in \text{St}(m)$ est tel que $u \in F(x)$.

Les deux dernières parties, $\text{LDA}_i^{\text{1er}}$ et $\text{LDA}_i^{\text{2nd}}$, contiennent les informations relatives aux deux (quasi-)représentants u_1 et u_2 de u sur : les bordures des deux panneaux 1-voisins $F(w_1)$ et $F(w_2)$ de $F(x)$, si $F(x)$ est un cône ; ou sur la bordure de $F(x)$ si $F(x)$ est un panneau. Si $F(x)$ est un cône, alors

- (1) $\text{LDA}_i^{\text{1er}[\text{Porte_LDT}]}(u) := \text{LD}_{\partial^*F(w_1)}(u_1)$, où $\text{LD}_{\partial^*F(w_1)}(u_1)$ est une étiquette de distance exacte de u_1 dans l'arbre étoilé $\partial^*F(w_1)$;
- (2) $\text{LDA}_i^{\text{1er}[\text{Dist}]}(u) := d_G(u, u_1)$.

Et si $F(x)$ est un panneau, alors

- (1) $\text{LDA}_i^{\text{1er[Porte_LDT]}}(u) := \text{LD}_{\partial^* F(x)}(u_1)$;
- (2) $\text{LDA}_i^{\text{1er[Dist]}}(u) := d_G(u, u_1)$.

La partie $\text{LDA}_i^{\text{2nd}}$ est d'une forme identique à $\text{LDA}_i^{\text{1er}}$ avec u_2 à la place de u_1 (et $\partial^* F(w_2)$ au lieu de $\partial^* F(w_1)$ si $F(x)$ est un cône).

Nous continuons avec le pseudo-code de l'algorithme d'encodage.

Algorithme 5 : ENC_DIST ($G, \text{LDA}(V)$)

Entrée : Un graphe ponté $G = (V, E)$ sans K_4 et un étiquetage $\text{LDA}(V)$ de ses sommets consistant, initialement, en un identifiant unique $\text{id}(u)$ donné à chaque sommet $u \in V$

- 1 si $V = \{v\}$ alors stop ;
- 2 Trouver un sommet médian m de G ;
- 3 $\text{L}_{\text{St}(m)}(\text{St}(m)) \leftarrow \text{Enc_Etoile_Pont}(\text{St}(m))$;
- 4 pour chaque panneau $F(x) \in \mathcal{F}_m$ faire
 - 5 $\text{LD}_{\partial^* F(x)}(\partial^* F(x)) \leftarrow \text{Enc_Dist_Arbre}(\partial^* F(x))$;
 - 6 pour chaque $u \in F(x)$ faire
 - 7 Trouver les représentants u_1 et u_2 de u sur $\partial^* F(x)$;
 - 8 $L_{\text{St}} \leftarrow (\text{id}(m), d_G(u, m), \text{L}_{\text{St}(m)}(x))$;
 - 9 $L_{\text{1er}} \leftarrow (\text{LD}_{\partial^* F(x)}(u_1), d_G(u, u_1))$;
 - 10 $L_{\text{2nd}} \leftarrow (\text{LD}_{\partial^* F(x)}(u_2), d_G(u, u_2))$;
 - 11 $\text{LDA}(u) \leftarrow \text{LDA}(u) \circ (L_{\text{St}}, L_{\text{1er}}, L_{\text{2nd}})$;
 - 12 **ENC_DIST** ($F(x), \text{LDA}(V)$) ;
- 13 pour chaque cône $F(x) \in \mathcal{F}_m$ faire
 - 14 pour chaque $u \in F(x)$ faire
 - 15 Trouver les panneaux $F(w_1)$ et $F(w_2)$ 1-voisins de $F(x)$;
 - 16 Choisir des représentants u_1^+ et u_2^+ de u sur $F(w_1)$ et $F(w_2)$;
 - 17 $L_{\text{St}} \leftarrow (\text{id}(m), d_G(u, m), \text{L}_{\text{St}(m)}(x))$;
 - 18 $L_{\text{1er}} \leftarrow (\text{LD}_{\partial^* F(w_1)}(u_1^+), d_G(u, u_1^+))$;
 - 19 $L_{\text{2nd}} \leftarrow (\text{LD}_{\partial^* F(w_2)}(u_2^+), d_G(u, u_2^+))$;
 - 20 $\text{LDA}(u) \leftarrow \text{LDA}(u) \circ (L_{\text{St}}, L_{\text{1er}}, L_{\text{2nd}})$;
 - 21 **ENC_DIST** ($F(x), \text{LDA}(V)$) ;

8.7.2 Requêtes de distances

De même que nous l'expliquions dans la section 7.7.1 du chapitre 7, étant données les étiquettes $\text{LDA}(u)$ et $\text{LDA}(v)$ de deux sommets, **DISTANCE** commence par déterminer le statut de la paire (u, v) . Pour cela, elle recherche le médian séparateur m de u et v , ou plutôt, la partie i d'étiquette correspondant à l'étape ayant fait émerger ce médian. Notons que la même astuce que celle présentée en section 7.7.3 nous permet de trouver ce médian en temps constant. Une fois cette étape effectuée, **DISTANCE** déterminera que deux sommets sont 1pc-voisins si et seulement l'identifiant (c.-à-d., l'étiquetage dans $\text{St}(m)$) de la fibre de l'un

est inclus dans celui de la fibre de l'autre. Dans ce cas, la fonction de décodage fera appel à la procédure `Dist_1pc-voisins` qui suit. Cette dernière, basée sur le lemme 54, suppose que son premier argument u est dans un panneau et son second v est dans un cône. Précisons aussi que nous supposons que $LDA_i^{1er}(v)$ contient toujours les informations pour se rendre dans le panneau dont l'identifiant est le minimum des deux valeurs identifiant le cône de v .

```

fonction Dist_1pc-voisins( $LDA_i(u)$ ,  $LDA_i(v)$ ):
  dir ← 1er ; // Si  $LDA_i^{St[racine]}(u) = \max\{j : j \in LDA_i^{St[racine]}(v)\}$ 
  si  $LDA_i^{St[racine]}(u) = \min\{j : j \in LDA_i^{St[racine]}(v)\}$  alors
    dir ← 2nd ;
   $v' \leftarrow LDA_i^{dir[rep\_LDT]}(v)$  ;
   $u_1, u_2 \leftarrow LDA_i^{1er[rep\_LDT]}(u)$ ,  $LDA_i^{2nd[rep\_LDT]}(u)$  ;
   $d_1, d_2 \leftarrow Dist\_Arbre(v', u_1)$ ,  $Dist\_Arbre(v', u_2)$  ;
  retourner
   $LDA_i^{dir[Dist]}(v) + \min\{d_1 + LDA_i^{1er[Dist]}(u), d_2 + LDA_i^{2nd[Dist]}(u)\}$  ;

```

En fait, dans tous les cas restant (c.-à-d., u et v ni proches, ni 1pc-voisins), l'algorithme de décodage retournera la somme $d_G(u, m) + d_G(v, m)$. Cette somme est comprise entre $d_G(u, v)$ et $2 \cdot d_G(u, v)$ d'après les lemmes 52 et 53.

Algorithme 6 : DISTANCE ($LDA(u)$, $LDA(v)$)

Entrée : Les étiquettes $LDA(u)$ et $LDA(v)$ de deux sommets u et v de G

Sortie : Une valeur entre $d_G(u, v)$ et $2 \cdot d_G(u, v)$

- 1 si $LDA_0(u) = LDA_0(v)$ /* $u = v$ */ **alors retourner** 0 ;
 - 2 Soit i le plus grand entier tel que $LDA_i^{St[Med]}(u) = LDA_i^{St[Med]}(v)$;
// Si u dans un panneau 1pc-voisin du cône de v
 - 3 si $LDA_i^{St[racine]}(u) \subsetneq LDA_i^{St[racine]}(v)$ **alors**
 - 4 | **retourner** `Dist_1pc-voisins`($LDA_i(u)$, $LDA_i(v)$) ;
// Si v dans un panneau 1pc-voisin du cône de u
 - 5 si $LDA_i^{St[racine]}(v) \subsetneq LDA_i^{St[racine]}(u)$ **alors**
 - 6 | **retourner** `Dist_1pc-voisins`($LDA_i(v)$, $LDA_i(u)$) ;
// Dans tous les autres cas
 - 7 **retourner** $LDA_i^{St[Dist]}(u) + LDA_i^{St[Dist]}(v)$;
-

Remarque 33. Si, pour deux sommets u et v , à l'étape i , la fonction de décodage détecte qu'ils sont dans un cône avec le même identifiant, il suffit de regarder s'ils stockent le même médian à l'étape $i + 1$. Si oui, c'est qu'ils étaient proches. Si non, c'est qu'ils sont dans des cônes 2cc-voisins. La distance exacte entre eux est alors $d_G(u, m) + d_G(v, m) - 2$. Dans ce cas, puisque nous autorisons une erreur multiplicative de 4, l'algorithme `DISTANCE` retourne simplement $d_G(u, m) + d_G(v, m)$.

8.7.3 Correction et complexité

De même que dans le chapitre précédent, nous supposons le graphe $G = (V, E)$ représenté par une matrice de distances. De nouveau, l'étoile s'encode en temps $O(\min\{\Delta^2, n\})$, où Δ dénote le degré maximum de G et n son nombre sommets. Les étapes les plus coûteuses, qui consistent en le calcul d'un sommet médian et de représentants (ou quasi-représentants), nécessitent un temps polynomial. Le nombre d'étapes récursives est borné par $O(\log n)$.

Comme dans l'analyse de la longueur des étiquettes du schéma de distance pour les graphes médians sans-cubes, nous avons ici $O(\log n)$ parties dans les étiquettes (du fait du lemme 47). Chaque partie stocke l'identifiant unique d'un médian, trois distances et un identifiant de fibre pour un coût de $O(\log n)$ bits chacun, ainsi que deux étiquettes données par un schéma de distance dans un arbre de $O(n)$ sommets, ce qui coûte $O(\log^2 n)$ bits supplémentaires. Au final, chaque étiquette nécessite $O(\log^3 n)$ bits.

De la même manière que nous l'avons indiqué dans le chapitre 7 (section 7.7.3), pour deux sommets u et v , étiquetés $LDA(u)$ et $LDA(v)$ par `ENC_DIST`, nous pouvons récupérer le médian séparateur m de u et v en temps constant au prix de concaténer à chaque étiquette $O(\log^2 n)$ bits correspondant à un étiquetage NCAD (dans l'arbre des séparateurs récursifs). De cela découle directement le fait que les requêtes de distances (approchées) soient satisfaites en temps $O(1)$.

Pour ce qui est de la correction de l'algorithme `DISTANCE` (c.-à-d., le fait qu'il retourne des distances avec une erreur multiplicative de 4 au maximum), elle découle directement des lemmes 52 et 53 pour les cas non 1pc-voisins, et du lemme 54 dans le cas où les sommets sont 1pc-voisins. Ceci conclut la preuve du théorème 22.

Remarque 34. De la même manière que nous avons adapté notre schéma de distance pour les graphes médians sans cube afin d'obtenir un schéma de routage, nous pouvons adapter notre schéma de distance approché pour les graphes pontés sans K_4 pour obtenir un schéma de routage approché.

Conclusion

Dans cette partie, nous avons présenté des schémas de distance et de routage pour les graphes médians sans cube avec n sommets, ainsi qu'un schéma de distance 4-approché pour les graphes pontés sans K_4 avec n sommets. Ces schémas utilisent tous des étiquettes de $O(\log^3 n)$ bits, attribuées en temps polynomial et décodables en temps constant. Ils sont basés sur une partition (appliquée récursivement) d'un graphe G en l'étoile d'un sommet médian et ses fibres. Dans les deux cas (des graphes médians et pontés), les fibres appartiennent toujours à la classe de graphes correspondante, et elles possèdent au plus la moitié des sommets de G . Nous avons ensuite distingué deux types de fibres : des cônes et des panneaux.

Dans le cas des graphes médians du chapitre 7, nous avons classifié les paires de sommets (u, v) de G en proches, 1-voisins, 2-voisins et éloignés. Lorsque u et v sont proches, $d_G(u, v)$ est calculée à une étape suivante de la récursion. Dans les autres cas, nous avons montré comment retrouver $d_G(u, v)$ en conservant dans les étiquettes de u et v des informations sur certaines portes/empreintes de ces sommets sur des ensembles appropriés. L'ingrédient principal faisant fonctionner nos schémas pour les graphes médians sans cube est le fait que les bordures des fibres soient des arbres isométriques quasi-portés. Malheureusement, cette dernière propriété des fibres est un obstacle pour la généralisation de notre schéma à des graphes médians de dimension quelconque car elle n'est plus vraie à partir des graphes médians de dimension 3. En dimension supérieure à 3, en plus du fait que les sommets des fibres puissent avoir un nombre quelconque d'empreintes sur leur bordures, un autre problème est dû au fait que cette bordure ne soit plus nécessairement un graphe médian. Cela rend impossible une application récursive de notre schéma à cette bordure (dont nous montrons qu'en revanche elle est bien de dimension $< \dim(G)$).

Nous pouvons rappeler ici que CHEPOI et HAGEN [56] ont montré qu'il existe une constante¹ M telle que tout graphe médian sans cube G de degré maximum Δ puisse être plongé dans le produit cartésien d'au plus $\varepsilon(\Delta) := M\Delta^{26}$ arbres. Cela implique directement que les graphes médians sans cube admettent un schéma de distance avec des étiquettes de longueur au plus $O(\varepsilon(\Delta) \log^2 n)$. Pour un degré maximum Δ supposé constant, un schéma basé sur ce résultat

1. La constante obtenue par les auteurs vaut 1165226.

est (asymptotiquement) plus “performant” que le schéma en $O(\log^3 n)$ que nous proposons dans le chapitre 7. Ceci dit, supposer fixé le degré maximum n’est pas tout à fait raisonnable, et nous pouvons en fait faire la remarque qui suit. Les graphes médians généraux avec un degré maximum Δ constant ont des cubes de taille bornée. Il s’en suit qu’une étoile $\text{St}(m)$ ne peut pas contenir plus de 2^Δ sommets. Puisque chaque fibre est portée et possède un nombre constant de fibres voisines, chaque sommet peut stocker sa porte vers chacune des fibres de $\text{St}(m)$, à chaque étape de la récursion. Cela conduit à des schémas de distance et de routage avec étiquettes de $O(2^\Delta \log^3 n)$ bits pour les graphes médians de degré maximum Δ .

Quoiqu’il en soit, le problème devient vraiment difficile pour une assez grande dimension puisque CHEPOI et HAGEN [56] fournissent aussi un exemple de graphe médian (en fait, de complexe cubique $\text{CAT}(0)$) de dimension 5 et de degrés bornés n’admettant pas de plongement isométrique dans le produit cartésien d’un nombre fini d’arbres. Cela explique, d’une certaine façon, la difficulté d’obtenir un schéma poly-logarithmique pour cette famille et justifie la question de l’existence d’un tel schéma. Toujours est-il que nous n’avons, pour l’instant, aucune raison de penser que les graphes médians n’admettent pas de schéma poly-logarithmique. Nous formulons donc la question suivante :

Question 3. Existe-t-il un schéma de distance (et/ou de routage) poly-logarithmique pour les graphes médians de dimension arbitraire (mais fixée) ?

Pour revenir au cas des graphes pontés du chapitre 8, nous y avons aussi classifié les paires de sommets (u, v) . Nous avons déterminé dans quel cas passer par le médian donne une distance exacte (c’est le cas des sommets éloignés), et dans quels cas cela peut conduire à une erreur multiplicative de 2 (sommets 1pp-, 2cc-, 2pc- et 3cc-voisins). Enfin, nous avons distingué un cas plus problématique (celui des sommets 1pc-voisins), dans lequel passer par le médian n’est pas satisfaisant (car l’erreur devient arbitraire), mais dans lequel nous pouvons faire une erreur multiplicative de 4 en conservant dans les étiquettes de u et v des informations relatives à certains représentants/quasi-représentants de ces sommets. Nous pouvons préciser que le schéma de distance approché que nous fournissons pour les graphes pontés sans K_4 peut être adapté (comme celui des graphes médians sans cube) pour obtenir un schéma de routage approché.

Contrairement au cas des graphes médians, nous n’avons pas défini les étoiles et les fibres dans le cas des graphes pontés généraux, mais seulement dans le cas des graphes pontés sans K_4 . En effet, une telle définition n’est plus aussi claire lorsque les cliques peuvent être de tailles et d’intersection arbitraires. Nous ne pouvons donc pas pointer aussi précisément que pour les médians les difficultés à la généralisation de notre schéma dans ce cas. Toujours est-il que la structure assez arbitraire des graphes pontés généraux laisse à penser que la problématique

de leur trouver un schéma de distance poly-logarithmique semble aussi compliquée, même en autorisant une erreur multiplicative ou additive (constante). Nous posons donc naturellement les questions suivantes :

Question 4. Existe-t-il un schéma de distance (exact ou avec erreur constante additive et/ou multiplicative) poly-logarithmique pour les graphes pontés de clique maximale arbitraire (mais fixée) ?

Une autre piste pour répondre aux questions 3 et 4 consisterait à fixer la δ -hyperbolicité, une autre propriété métrique importante, et structurante. Par exemple, les graphes pontés sans C_6 induits (c.-à-d., le 1-squelette des complexes systoliques 7-larges) sont d'hyperbolicité constante (JANUSZKIEWICZ et ŚWIĄTKOWSKI [104] montrent que ces graphes sont 11-hyperboliques, puis CHEPOI et al. [54] prouvent qu'ils sont 1-hyperboliques). Dans les graphes médians, la δ -hyperbolicité dépend de la taille des grilles [54]. Or, la présence d'un hypercube de grande dimension implique celle d'une grande grille. Fixer la δ -hyperbolicité borne donc la dimension. Dans les graphes pontés, la δ -hyperbolicité dépend des triangles métriques [54]. En fait, fixer la δ -hyperbolicité semble un premier pas naturel et, si les réponses aux questions 3 et 4 s'avéraient négatives, elles pourraient être vérifiées dans le cas de graphes médians ou pontés de δ -hyperbolicité bornée. La question qui suit présente donc déjà un grand intérêt.

Question 5. Existe-t-il des schémas de distance (ou de routage) pour les familles des graphes médians ou pontés de δ -hyperbolicité bornée ?

Notons que GAVOILLE et LY [82] ont établi que les graphes quelconque de δ -hyperbolicité bornée n'admettent un schéma de distance poly-logarithmique qu'à condition d'autoriser une erreur multiplicative d'au moins $\Omega(\log \log n)$. Ce résultat n'est toutefois pas en contradiction avec la question 5 qui se pose pour des familles avec plus de propriétés métriques.

Remarquons pour finir que notre schéma pour les graphes médians pourrait bien s'adapter aux graphes quasi-médians de *dimension cubique 2* (c.-à-d., ne contenant pas de sous-graphe isomorphe à un produit cartésien de plus de deux cliques), au moins pour donner un schéma de distance avec erreur additive constante. En effet, ces graphes possèdent de nombreuses propriétés nécessaires au bon fonctionnement de notre schéma telles que l'équivalence entre le fait d'être porté et celui d'être convexe (hérité des graphes faiblement modulaires), ou le fait que les étoiles (c.-à-d., "bouquets de graphes de Hamming") et les fibres soient portées (voir la thèse de GENEVOIS [88] pour une partie de ces propriétés). Évidemment, il reste à étudier, entre autre, l'"interaction" entre ces fibres. En particulier, contrairement au cas des graphes médians (et de façon similaire à celui des pontés), des panneaux peuvent, dans ce cas, être voisins d'autres panneaux car le voisinage d'un sommet contient des cliques.

Conclusion et perspectives

Dans la partie II de cette thèse, nous nous sommes intéressés à des généralisations du théorème de HAUSSLER, LITTLESTONE et WARMUTH [100] concernant la densité des sous-graphes d'hypercubes. Nous avons en premier lieu étendu ce résultat pour les sous-graphes de produits cartésiens de graphes. Nous avons montré que *si G est un sous-graphe d'un produit cartésien $G_1 \square \dots \square G_m$, alors $\text{dens}(G) \leq \max\{\text{deg}^*(G_i) : 1 \leq i \leq m\} \cdot \log |V(G)|$* . Nous avons ensuite amélioré ce résultat en étendant la notion classique de VC-dimension pour les produits cartésiens de graphes quelconques. Nous avons alors prouvé que, *si G est un sous-graphe d'un produit cartésien Γ de m graphes excluant un mineur fixé H , alors $\text{dens}(G) \leq \mu(H) \cdot \text{VC-dim}^*(G)$* . En nous basant sur ces résultats, ainsi que sur le théorème 9 de NASH-WILLIAMS [121] et la proposition 6 de KANNAN, NAOR et RUDICH [106], nous avons déduit des bornes supérieures sur la longueur des étiquettes attribuées par un schéma d'adjacence sur les sous-graphes de produits cartésiens de graphes de mineur exclu H , cordaux, démontables, de dégénérescence bornée, ou d'octaèdres. Nous avons aussi introduit un paramètre de VC-densité dans le but d'améliorer notre borne supérieure basée sur la VC-dimension, mais nous ne sommes pas parvenus à l'utiliser assez précisément.

Comme nous l'avons vu dans le chapitre 2, des généralisations ou extensions du lemme 2 de Sauer ont fait l'objet de plusieurs travaux. En particulier, ce lemme a été étendu aux graphes de Hamming par HAUSSLER et LONG [101], puis aux puissances cartésiennes d'un graphe par CESA-BIANCHI et HAUSSLER [41]. Nous pourrions donc imaginer établir un analogue dans le cas plus général des produits cartésiens de graphes quelconques en utilisant notre notion étendue de VC-dimension. La même question s'étend naturellement à une généralisation du lemme 3 du sandwich.

Dans un second chapitre de la partie II, nous avons considéré des sous-graphes de demi-cubes. Nous avons défini la notion de clique-VC-dimension, adaptée à l'étude de leur densité. Nous avons alors montré que *si G est un sous-graphe de demi-cube induit par une famille d'ensemble de clique-VC-dimension d , alors $\text{dens}(G) \leq \binom{d}{2}$* . Nous avons remarqué que les demi-cubes contenaient des classes de graphes dont la densité dépendait essentiellement de deux paramètres : la taille de la clique maximale (cliques, graphes de Johnson, etc.) ; et la VC-dimension (hypercubes). Nous posons donc la question de l'existence d'une fonction

de ces deux paramètres majorant la densité des sous-graphes de demi-cubes.

Dans la partie III, nous avons établi des schémas de distance et de routage pour les familles des graphes médians sans cube et des graphes pontés sans K_4 . Chacun de ces schémas utilise des étiquettes de $O(\log^3)$ bits, décodées en temps constant. Ces étiquettes sont construites en utilisant une partition des graphes en étoiles et en fibres. En ce qui concerne les graphes médians, les réponses calculées pour la distance et le routage sont exactes. Pour les graphes pontés, une erreur multiplicative de 4 peut être faite sur le calcul des distances.

Outre des généralisations de nos schémas à des graphes médians de dimension quelconque ou à des graphes pontés de clique-maximale quelconque, nous pouvons envisager la piste de recherche suivante. Comme nous en avons déjà fait la remarque, les graphes pontés sans K_4 et les graphes médians sans cubes sont des généralisations respectives d'une partie des $(3, 6)$ - et $(4, 4)$ -graphes planaires étudiés par CHEPOI, DRAGAN et VAXÈS [53]. En fait, ces auteurs donnent un schéma générique valable pour les $(3, 6)$ -, $(6, 3)$ - et $(4, 4)$ -graphes planaires. La question est donc de savoir si nous pouvons obtenir des schémas semblables regroupant leurs versions non-planaires. Plus précisément, les graphes pontés (sans K_4) et médians (sans Q_3) que nous avons étudiés sont aussi le 1-squelette de deux complexes $CAT(0)$ (systolique et cubique) 2-dimensionnels issus de la *théorie des petites simplifications* (voir le livre de LYNDON et SCHUPP [115]). Nous pourrions donc nous intéresser au 1-squelette de (m, n) -complexes $CAT(0)$ (complexes plus généraux, eux-mêmes issus de la théorie des petites simplifications). Où un (m, n) -complexe est un complexe 2-dimensionnel dans lequel les cycles sont de longueur au moins m et les sommets sont de degré au moins n .

Nous posons aussi la question de la portée de notre technique basée sur une partition en étoiles et en fibres. Dans quelles autres familles pouvons-nous définir une telle partition, avec des propriétés suffisantes, comme par exemple : fibres demeurant dans la classe considérée ; bordures isométrique/portée/quasi-portée/étoilée/etc. ; peu de classes de paires de sommets (u, v) “problématiques” (ce terme étant à définir, par exemple par *il n'existe pas de plus court (u, v) -chemin passant par l'étoile*) ; etc. Comme nous l'avons conjecturé dans la conclusion de la partie III, ce type de partition pourrait être utilisé dans les graphes quasi-médians de dimension cubique 2.

Pour finir, nous devons remarquer que nous n'avons pas considéré les problèmes de bornes inférieures dans nos résultats. Ces problèmes sont généralement difficiles et utilisent des méthodes et techniques très différentes. Ils constituent donc une problématique qu'il serait intéressant de traiter. Une première question concernerait les graphes pontés sans K_4 : un schéma poly-logarithmique implique-t-il nécessairement une erreur (multiplicative ou additive) ? Pourrions-nous faire mieux, dans un premier temps, qu'un facteur 4 en exhibant des propriétés supplémentaires sur les fibres ?

Bibliographie

- [1] Ittai ABRAHAM, Daniel DELLING, Amos FIAT, Andrew V. GOLDBERG et Renato F. WERNECK. « VC-dimension and shortest path algorithms ». In : *ICALP*. Springer. 2011, p. 690-699 (cf. p. [10](#), [44](#)).
- [2] Ittai ABRAHAM, Daniel DELLING, Andrew V GOLDBERG et Renato F WERNECK. « Hierarchical hub labelings for shortest paths ». In : *ESA*. Springer. 2012, p. 24-35 (cf. p. [12](#)).
- [3] Ittai ABRAHAM et Cyril GAVOILLE. « On approximate distance labels and routing schemes with affine stretch ». In : *DISC*. Springer. 2011, p. 404-415 (cf. p. [52](#), [55](#)).
- [4] Aaron ABRAMS et Robert GHRIST. « State complexes for metamorphic robots ». In : *Intl. J. Robotics Res.* 23 (2004), p. 811-826 (cf. p. [63](#)).
- [5] Foto AFRATI, Christos H. PAPADIMITRIOU et George PAPAGEORGIOU. « The complexity of cubical graphs ». In : *Information and control* 66 (1985), p. 53-60 (cf. p. [38](#)).
- [6] Rudolf AHLWEDE et Ning CAI. « A counterexample to Kleitman's conjecture concerning an edge-isoperimetric problem ». In : *Combinatorics, Probability and Computing* 8 (1999), p. 301-305 (cf. p. [43](#)).
- [7] Noga ALON et Michael TARSI. « Colorings and orientations of graphs ». In : *Combinatorica* 12 (1992), p. 125-134 (cf. p. [88](#)).
- [8] Stephen ALSTRUP, Philip BILLE et Theis RAUHE. « Labeling Schemes for Small Distances in Trees ». In : *SODA*. SIAM, 2003, p. 689-698 (cf. p. [53](#)).
- [9] Stephen ALSTRUP, Søren DAHLGAARD et Mathias Bæk Tejs KNUDSEN. « Optimal induced universal graphs and adjacency labeling for trees ». In : *Journal of the ACM (JACM)* 64 (2017), p. 27 (cf. p. [48](#)).
- [10] Stephen ALSTRUP, Søren DAHLGAARD, Mathias Bæk Tejs KNUDSEN et Ely PORAT. « Sublinear Distance Labeling ». In : *ESA*. 2016, 5 :1-5 :15 (cf. p. [52](#)).
- [11] Stephen ALSTRUP, Cyril GAVOILLE, Esben Bstrup HALVORSEN et Holger PETERSEN. « Simpler, faster and shorter labels for distances in graphs ». In : *SODA*. 2016, p. 338-350 (cf. p. [51](#)).

-
- [12] Stephen ALSTRUP, Cyril GAVOILLE, Haim KAPLAN et Theis RAUHE. « Nearest common ancestors : a survey and a new distributed algorithm ». In : *SPAA*. ACM. 2002, p. 258-264 (cf. p. 11).
- [13] Stephen ALSTRUP, Inge Li GØRTZ, Esben Bistrup HALVORSEN et Ely PORAT. « Distance Labeling Schemes for Trees ». In : *ICALP*. 2016, 132 :1-132 :16 (cf. p. 12, 51).
- [14] Stephen ALSTRUP, Haim KAPLAN, Mikkel THORUP et Uri ZWICK. « Adjacency Labeling Schemes and Induced-Universal Graphs ». In : *SIAM J. Discr. Math.* 33.1 (2019), p. 116-137 (cf. p. 50, 51).
- [15] Stephen ALSTRUP et Theis RAUHE. « Small induced-universal graphs and compact implicit graph representations ». In : *FOCS*. IEEE. 2002, p. 53-62 (cf. p. 48, 50).
- [16] Richard P ANSTEE et Martin FARBER. « On bridged graphs and cop-win graphs ». In : *J. Comb. Theory, Series B* 44.1 (1988), p. 22-28 (cf. p. 66).
- [17] Martin ANTHONY, Graham BRIGHTWELL et Colin COOPER. « The Vapnik-Chervonenkis dimension of a random graph ». In : *Discr. Math.* 138 (1995), p. 43-56 (cf. p. 44).
- [18] S.P. AVANN. « Metric ternary distributive semi-lattices ». In : *Proc. Amer. Math. Soc.* 12 (1961), p. 407-414 (cf. p. 58).
- [19] Hans-Jürgen BANDELDT. « Retracts of hypercubes ». In : *J. Graph Theory* 8 (1984), p. 501-510 (cf. p. 60).
- [20] Hans-Jürgen BANDELDT. « Hereditary modular graphs ». In : *Combinatorica* 8.2 (1988), p. 149-157 (cf. p. 59).
- [21] Hans-Jürgen BANDELDT, Victor CHEPOI, Andreas DRESS et Jack KOOLEN. « Combinatorics of lopsided sets ». In : *European J. Combin.* 27 (2006), p. 669-689 (cf. p. 10, 39).
- [22] Hans-Jürgen BANDELDT et Victor CHEPOI. « A Helly theorem in weakly modular space ». In : *Discr. Math.* 160.1-3 (1996), p. 25-39 (cf. p. 56).
- [23] Hans-Jürgen BANDELDT et Victor CHEPOI. « Metric graph theory and geometry : a survey ». In : *Contemporary Mathematics* 453 (2008), p. 49-86 (cf. p. 31, 58, 60, 61, 71).
- [24] Hans-Jürgen BANDELDT, Victor CHEPOI et David EPPSTEIN. « Ramified rectilinear polygons : coordinatization by dendrons ». In : *Discr. Comput. Geom.* 54 (2015), p. 771-797 (cf. p. 115).
- [25] Hans-Jürgen BANDELDT et Marcel van DE VEL. « Embedding topological median algebras in products of dendrons ». In : *Proc. London Math. Soc.* s3-58 (1989), p. 439-453 (cf. p. 60, 119).
- [26] Jean-Pierre BARTHÉLEMY et Julien CONSTANTIN. « Median graphs, parallelism and posets ». In : *Discr. Math.* 111 (1993), p. 49-63 (cf. p. 27).

- [27] Fabrice BAZZARO et Cyril GAVOILLE. « Localized and compact data-structure for comparability graphs ». In : *Discr. Math.* 309 (2009), p. 3465-3484 (cf. p. 51).
- [28] Sergei L. BEZRUKOV. « Edge isoperimetric problems on graphs ». In : *Graph Theory and Combinatorial Biology* 7 (1999), p. 157-197 (cf. p. 41-43).
- [29] Sandeep N BHATT et Charles E LEISERSON. « How to assemble tree machines ». In : *STOC*. ACM. 1982, p. 77-84 (cf. p. 49).
- [30] Louis J. BILLERA, Susan P. HOLMES et Karen VOGTMANN. « Geometry of the space of phylogenetic trees ». In : *Adv. Appl. Math.* 27 (2001), p. 733-767 (cf. p. 63).
- [31] Garrett BIRKHOFF et Stephen A. KISS. « A ternary operation in distributive lattices ». In : *Bull. Amer. Math. Soc.* 53 (1947), p. 745-752 (cf. p. 58).
- [32] Béla BOLLOBÁS, Don COPPERSMITH et Michael ELKIN. « Sparse distance preservers and additive spanners ». In : *SIAM J. Discr. Math.* 19 (2005), p. 1029-1055 (cf. p. 52).
- [33] Béla BOLLOBÁS, I LEADER et AJ RADCLIFFE. « Reverse kleitman inequalities ». In : *Proceedings of the London Mathematical Society* 3.1 (1989), p. 153-168 (cf. p. 10, 39).
- [34] Béla BOLLOBÁS et Andrew J. RADCLIFFE. « Defect sauer results ». In : *J. Comb. Theory, Series A* 72 (1995), p. 189-208 (cf. p. 39).
- [35] Stéphane BOUCHERON. « Théorie de l'apprentissage ». In : *Hermes, Paris* (1992) (cf. p. 35).
- [36] Nicolas BOUSQUET et Stéphan THOMASSÉ. « VC-dimension and Erdős–Pósa property ». In : *Discr. Math.* 338 (2015), p. 2302-2317 (cf. p. 10, 44, 98).
- [37] Boštjan BREŠAR, Jérémie CHALOPIN, Victor CHEPOI, Tanja GOLOGRANC et Damian OSAJDA. « Bucolic complexes ». In : *Advances in Mathematics* 243 (2013), p. 127-167 (cf. p. 90).
- [38] Melvin A. BREUER. « Coding the vertexes of a graph ». In : *Trans. Inf. Theory* 12.2 (1966), p. 148-153 (cf. p. 10, 47).
- [39] Melvin A. BREUER et Jon FOLKMAN. « An unexpected result in coding the vertices of a graph ». In : *J. Math. Anal. Appl.* 20 (1967), p. 583-600 (cf. p. 114).
- [40] Martin R BRIDSON et André HAEFLIGER. *Metric Spaces of Non-Positive Curvature*. T. 319. Springer Science & Business Media, 2013 (cf. p. 62).
- [41] Nicolö CESA-BIANCHI et David HAUSSLER. « A graph-theoretic generalization of the Sauer-Shelah lemma ». In : *Discrete Appl. Math.* 86 (1998), p. 27-35 (cf. p. 10, 44, 77, 180).

-
- [42] Jérémie CHALOPIN et Victor CHEPOI. « A Counterexample to Thiagarajan’s Conjecture on Regular Event Structures ». In : *ICALP*. arXiv :1605.08288, 2017, 101 :1-101 :14 (cf. p. 61).
- [43] Jérémie CHALOPIN, Victor CHEPOI, Hiroshi HIRAI et Damian OSAJDA. « Weakly modular graphs and nonpositive curvature ». In : *arXiv preprint arXiv :1409.3892* (2014) (cf. p. 57).
- [44] Jérémie CHALOPIN, Victor CHEPOI, Shay MORAN et Manfred K WARMUTH. « Unlabeled sample compression schemes and corner peelings for ample and maximum classes ». In : *arXiv preprint arXiv :1812.02099* (2018) (cf. p. 40).
- [45] Marc CHASTAND. « Fiber-complemented graphs—I : structure and invariant subgraphs ». In : *Discr. Math.* 226 (2001), p. 107-141 (cf. p. 122).
- [46] Victor CHEPOI. « Isometric subgraphs of Hamming graphs and d-convexity ». In : *Cybernetics and Systems Analysis* 24.1 (1988), p. 6-11 (cf. p. 28).
- [47] Victor CHEPOI. « Classification of graphs by means of metric triangles ». Russian. In : *Metody Diskret. Analiz.* (1989), p. 75-93, 96 (cf. p. 56, 58, 59, 65).
- [48] Victor CHEPOI. « Bridged graphs are cop-win graphs : an algorithmic proof ». In : *J. Comb. Theory, Series B* 69.1 (1997), p. 97-100 (cf. p. 66).
- [49] Victor CHEPOI. « Graphs of some CAT(0) complexes ». In : *Adv. Appl. Math.* 24 (2000), p. 125-179 (cf. p. 12, 13, 63, 66).
- [50] Victor CHEPOI. « Basis graphs of even Delta-matroids ». In : *J. Comb. Theory, Series B* 97 (2007), p. 175-192 (cf. p. 31).
- [51] Victor CHEPOI. « Distance-preserving subgraphs of Johnson graphs ». In : *Combinatorica* 37 (2017), p. 1039-1055 (cf. p. 32).
- [52] Victor CHEPOI, Feodor F. DRAGAN, Bertrand ESTELLON, Michel HABIB, Yann VAXÈS et Yang XIANG. « Additive spanners and distance and routing labeling schemes for hyperbolic graphs ». In : *Algorithmica* 62 (2012), p. 713-732 (cf. p. 52).
- [53] Victor CHEPOI, Feodor F. DRAGAN et Yann VAXÈS. « Distance and routing labeling schemes for non-positively curved plane graphs ». In : *J. Algorithms* 61 (2006), p. 60-88 (cf. p. 11, 51, 55, 64, 67, 115, 130, 181).
- [54] Victor CHEPOI, Feodor DRAGAN, Bertrand ESTELLON, Michel HABIB et Yann VAXÈS. « Diameters, centers, and approximating trees of delta-hyperbolic geodesic spaces and graphs ». In : *SoCG*. ACM. 2008, p. 59-68 (cf. p. 179).
- [55] Victor CHEPOI, Bertrand ESTELLON et Yann VAXÈS. « Covering planar graphs with a fixed number of balls ». In : *Discrete & Computational Geometry* 37 (2007), p. 237-244 (cf. p. 44).

- [56] Victor CHEPOI et Mark F. HAGEN. « On embeddings of CAT(0) cube complexes into products of trees via colouring their hyperplanes ». In : *J. Comb. Theory, Ser. B* 103 (2013), p. 428-467 (cf. p. [177](#), [178](#)).
- [57] Victor CHEPOI, Arnaud LABOUREL et Sébastien RATEL. « On density of subgraphs of Cartesian products ». In : *J. Graph Theory (to appear)* abs/1711.11485 (2017) (cf. p. [11](#)).
- [58] Victor CHEPOI, Arnaud LABOUREL et Sébastien RATEL. « Distance labeling schemes for cube-free median graphs ». In : *MFCS*. T. 138. 2019, 15 :1-15 :14 (cf. p. [13](#)).
- [59] Victor CHEPOI, Arnaud LABOUREL et Sébastien RATEL. « On density of subgraphs of halved cubes ». In : *European J. Combin.* 80 (2019), p. 57-70 (cf. p. [11](#), [109](#)).
- [60] Victor CHEPOI et Daniela MAFTULEAC. « Shortest path problem in rectangular complexes of global nonpositive curvature ». In : *Comput. Geom.* 46 (2013), p. 51-64 (cf. p. [64](#)).
- [61] Victor CHEPOI et Damian OSAJDA. « Dismantlability of weakly systolic complexes and applications ». In : *Transactions of the American Mathematical Society* 367.2 (2015), p. 1247-1272 (cf. p. [66](#)).
- [62] Fan RK CHUNG. « Universal graphs and induced-universal graphs ». In : *J. Graph Theory* 14.4 (1990), p. 443-454 (cf. p. [48](#), [49](#)).
- [63] Edith COHEN, Eran HALPERIN, Haim KAPLAN et Uri ZWICK. « Reachability and distance queries via 2-hop labels ». In : *SIAM J. Comput.* 32.5 (2003), p. 1338-1355 (cf. p. [12](#)).
- [64] Bruno COURCELLE et Rémi VANICAT. « Query efficient implementation of graphs of bounded clique-width ». In : *Discrete Appl. Math.* 131 (2003), p. 129-150 (cf. p. [51](#)).
- [65] Lenore J. COWEN. « Compact routing with minimum stretch ». In : *J. Algorithms* 38 (2001), p. 170-183 (cf. p. [55](#)).
- [66] Michel Marie DEZA et Monique LAURENT. *Geometry of Cuts and Metrics*. T. 15. Springer, 2009 (cf. p. [17](#), [29](#), [32](#), [91](#)).
- [67] Víctor DIEGO, Oriol SERRA et Lluís VENA. « On a problem by Shapozenko on Johnson graphs ». In : *Graphs Combin.* 34.5 (2018), p. 947-964 (cf. p. [43](#)).
- [68] Reinhard DIESTEL. *Graph Theory*. 1997 (cf. p. [17](#), [76](#)).
- [69] Dragomir DJOKOVIĆ. « Distance-preserving subgraphs of hypercubes ». In : *J. Comb. Theory, Series B* 14 (1973), p. 263-267 (cf. p. [26](#), [27](#), [32](#)).
- [70] Feodor F. DRAGAN et Chenyu YAN. « Collective tree spanners in graphs with bounded parameters ». In : *Algorithmica* 57 (2010), p. 22-43 (cf. p. [54](#), [55](#)).

-
- [71] A DRESS. « Towards a theory of holistic clustering ». In : *DIMACS Ser. Discrete Math. Theoret. Comput. Sci* 37 (1997), p. 271-289 (cf. p. 39).
- [72] Tamar EILAM, Cyril GAVOILLE et David PELEG. « Compact routing schemes with low stretch factor ». In : *J. Algorithms* 46 (2003), p. 97-114 (cf. p. 55).
- [73] Martin FARBER et Robert E. JAMISON. « On local convexity in graphs ». In : *Discr. Math.* 66.3 (1987), p. 231-247 (cf. p. 64).
- [74] Pierre FRAIGNIAUD et Cyril GAVOILLE. « Routing in trees ». In : *ICALP*. Springer. 2001, p. 757-772 (cf. p. 54).
- [75] Ofer FREEDMAN, Paweł GAWRYCHOWSKI, Patrick K. NICHOLSON et Oren WEIMANN. « Optimal distance labeling schemes for trees ». In : *PODC*. ACM. 2017, p. 185-194 (cf. p. 51, 53, 135, 140).
- [76] Zoltán FÜREDI et János PACH. « Traces of finite sets : extremal problems and geometric applications ». In : *Extremal problems for finite sets* 3 (1991), p. 255-282 (cf. p. 34).
- [77] Harold N GABOW et Herbert H WESTERMANN. « Forests, frames, and games : algorithms for matroid sums and applications ». In : *Algorithmica* 7.1-6 (1992), p. 465 (cf. p. 48).
- [78] Michael R. GAREY et Ronald L. GRAHAM. « On cubical graphs ». In : *J. Comb. Theory, Series B* 18 (1975), p. 84-95 (cf. p. 31).
- [79] Cyril GAVOILLE, Michal KATZ, Nir A. KATZ, Christophe PAUL et David PELEG. « Approximate distance labeling schemes ». In : *ESA*. Springer. 2001, p. 476-487 (cf. p. 52).
- [80] Cyril GAVOILLE et Arnaud LABOUREL. « Distributed Relationship Schemes for Trees ». In : *ISAAC*. 2007, p. 728-738 (cf. p. 53).
- [81] Cyril GAVOILLE et Arnaud LABOUREL. « Shorter implicit representation for planar graphs and bounded treewidth graphs ». In : *ESA*. Springer. 2007, p. 582-593 (cf. p. 49).
- [82] Cyril GAVOILLE et Olivier LY. « Distance labeling in hyperbolic graphs ». In : *ISAAC*. Springer. 2005, p. 1071-1079 (cf. p. 52, 179).
- [83] Cyril GAVOILLE et Christophe PAUL. « Distance labeling scheme and split decomposition ». In : *Discr. Math.* 273 (2003), p. 115-130 (cf. p. 51).
- [84] Cyril GAVOILLE et Christophe PAUL. « Optimal distance labeling for interval graphs and related graph families ». In : *SIAM J. Discr. Math.* 22 (2008), p. 1239-1258 (cf. p. 51).
- [85] Cyril GAVOILLE, David PELEG, Stéphane PÉRENNÈS et Ran RAZ. « Distance labeling in graphs ». In : *J. Algorithms* 53 (2004), p. 85-112 (cf. p. 11, 12, 51, 114).

- [86] Cyril GAVOILLE et Stéphane PÉRENNÈS. « Memory requirement for routing in distributed networks ». In : *PODC*. ACM, 1996, p. 125-133 (cf. p. 54).
- [87] Paweł GAWRYCHOWSKI et Przemysław UZNANSKI. « A note on distance labeling in planar graphs ». In : *CoRR* abs/1611.06529 (2016) (cf. p. 51).
- [88] Anthony GENEVOIS. « Géométrie cubique des graphes quasi-médians et applications à la théorie géométrique des groupes ». In : (2017) (cf. p. 179).
- [89] Robert GHIRST et Valerie PETERSON. « The geometry and topology of reconfiguration ». In : *Adv. Appl. Math.* 38 (2007), p. 302-323 (cf. p. 63).
- [90] Ronald L. GRAHAM. « On primitive graphs and optimal vertex assignments ». In : *Annals of the New York academy of sciences* 175 (1970), p. 170-186 (cf. p. 25).
- [91] Ronald L. GRAHAM et Peter M. WINKLER. « On isometric embeddings of graphs ». In : *Transactions of the American mathematical Society* 288 (1985), p. 527-536 (cf. p. 16, 29, 71).
- [92] Ronald L. GRAHAM et Henry O. POLLAK. « On embedding graphs in squashed cubes ». In : *Graph theory and applications*. Springer, 1972, p. 99-110 (cf. p. 11).
- [93] Mikhael GROMOV. « Hyperbolic groups ». In : *Essays in group theory*. Sous la dir. de S.M. GERSTEN. T. 8. Math. Sci. Res. Inst. Publ. Springer, New York, 1987, p. 75-263 (cf. p. 62, 63).
- [94] Frédéric HAGLUND. *Complexes simpliciaux hyperboliques de grande dimension*. Université de Paris-Sud. Département de Mathématique, 2003 (cf. p. 13, 66).
- [95] Richard H. HAMMACK, Wilfried IMRICH et Sandi KLAVŽAR. *Handbook of Product Graphs*. 2nd. CRC Press, Inc., 2011 (cf. p. 17, 25, 28, 29).
- [96] Steve HANNEKE et Liu YANG. « Minimax analysis of active learning ». In : *J. Mach. Learn. Res.* 16 (2015), p. 3487-3602 (cf. p. 109).
- [97] Lawrence H. HARPER. « Optimal assignments of numbers to vertices ». In : *Journal of the Society for Industrial and Applied Mathematics* 12 (1964), p. 131-135 (cf. p. 41-43).
- [98] Lawrence H. HARPER. *Global Methods for Combinatorial Isoperimetric Problems*. T. 90. Cambridge University Press, 2004 (cf. p. 42, 43, 91).
- [99] David HAUSSLER. « Sphere packing numbers for subsets of the Boolean n -cube with bounded Vapnik-Chervonenkis dimension ». In : *J. Comb. Theory, Series A* 69 (1995), p. 217-232 (cf. p. 31, 41, 42, 71, 94, 95, 102, 111, 112).
- [100] David HAUSSLER, Nick LITTLESTONE et Manfred K. WARMUTH. « Predicting $\{0, 1\}$ -functions on randomly drawn points ». In : *Information and Computation* 115 (1994), p. 248-292 (cf. p. 10, 14, 16, 31, 39-41, 70, 80, 111, 112, 180).

-
- [101] David HAUSSLER et Philip M. LONG. « A generalization of Sauer's lemma ». In : *J. Comb. Theory, Series A* 71 (1995), p. 219-240 (cf. p. [10](#), [44](#), [180](#)).
- [102] David HAUSSLER et Emo WELZL. « ε -nets and simplex range queries ». In : *Discrete & Computational Geometry* 2 (1987), p. 127-151 (cf. p. [10](#), [36](#), [44](#)).
- [103] Koyo HAYASHI. « A Polynomial Time Algorithm to Compute Geodesics in CAT(0) Cubical Complexes ». In : *ICALP*. 2018, 78 :1-78 :14 (cf. p. [64](#)).
- [104] Tadeusz JANUSZKIEWICZ et Jacek ŚWIĄTKOWSKI. « Simplicial nonpositive curvature ». In : *Publications Mathématiques de l'Institut des Hautes Études Scientifiques* 104.1 (2006), p. 1-85 (cf. p. [13](#), [66](#), [179](#)).
- [105] Camille JORDAN. « Sur les assemblages de lignes ». In : *J. Reine Angew. Math* 70.185 (1869), p. 81 (cf. p. [53](#)).
- [106] Sampath KANNAN, Moni NAOR et Steven RUDICH. « Implicit representation of graphs ». In : *SIAM J. Discr. Math.* 5 (1992), p. 596-603 (cf. p. [9](#), [10](#), [47](#), [48](#), [50](#), [70](#), [93](#), [114](#), [180](#)).
- [107] Haim KAPLAN et Tova MILO. « Short and Simple Labels for Small Distances and Other Functions ». In : *WADS*. Springer, 2001, p. 246-257 (cf. p. [52](#)).
- [108] Sandi KLAUŽAR et Henry Martyn MULDER. « Median Graphs : characterizations, Location Theory and Related Structures ». In : *J. Combin. Math. Combin. Comput.* 30 (1999), p. 103-127 (cf. p. [58](#)).
- [109] Donald Ervin KNUTH. *The Art of Computer Programming : Vol. 4. Fascicle 0, Introduction to Combinatorial Algorithms and Boolean Functions*. English. Boston, Mass. ; London : Addison-Wesley, 2008 (cf. p. [58](#), [61](#)).
- [110] Adrian KOSOWSKI et Laurent VIENNOT. « Beyond highway dimension : small distance labels using tree skeletons ». In : *SODA*. SIAM. 2017, p. 1462-1478 (cf. p. [52](#)).
- [111] Evangelos KRANAKIS, Danny KRIZANC, Berthold RUF, Jorge URRUTIA et Gerhard WÖGINGER. « The VC-dimension of set systems defined by graphs ». In : *Discrete Appl. Math.* 77 (1997), p. 237-257 (cf. p. [44](#)).
- [112] Dima KUZMIN et Manfred K. WARMUTH. « Unlabeled compression schemes for maximum classes ». In : *J. Mach. Learn. Res.* 8 (2007), p. 2047-2081 (cf. p. [40](#)).
- [113] James LAWRENCE. « Lopsided sets and orthant-intersection by convex sets ». In : *Pac. J. Math.* 104 (1983), p. 155-173 (cf. p. [10](#), [39](#)).
- [114] Nick LITTLESTONE et Manfred WARMUTH. « Relating data compression and learnability ». In : *Technical report, University of California, Santa Cruz* (1986) (cf. p. [40](#)).
- [115] Roger C LYNDON et Paul E SCHUPP. *Combinatorial Group Theory*. Springer, 2015 (cf. p. [67](#), [181](#)).

- [116] Stephen B. MAURER. « Matroid basis graphs. I ». In : *J. Comb. Theory, Series B* 14 (1973), p. 216-240 (cf. p. 31).
- [117] JW MOON. « On minimal n-universal graphs ». In : *Glasgow Mathematical Journal* 7.1 (1965), p. 32-33 (cf. p. 51).
- [118] Henry Martyn MULDER. *The Interval Function of a Graph*. T. 132. Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam, 1980 (cf. p. 60, 123).
- [119] Henry Martyn MULDER et Alexander SCHRIJVER. « Median graphs and Helly hypergraphs ». In : *Discr. Math.* 25 (1979), p. 41-50 (cf. p. 61).
- [120] John Harold MULLER. « Local structure in graph classes ». In : *Georgia Institute of Technology* (1988) (cf. p. 10, 47).
- [121] Crispin St. John Alvah NASH-WILLIAMS. « Edge-disjoint spanning trees of finite graphs ». In : *Journal of the London Mathematical Society* 1 (1961), p. 445-450 (cf. p. 10, 47, 48, 70, 93, 180).
- [122] Balas K NATARAJAN. *Machine learning : a theoretical approach*. Elsevier, 2014 (cf. p. 35).
- [123] Balas K. NATARAJAN. « On learning sets and functions ». In : *Machine Learning* 4 (1989), p. 67-97 (cf. p. 10, 44).
- [124] Ladislav NEBESKÝ. « Median graphs ». In : *Comm. Math. Univ. Carolinae* 12 (1971), p. 317-325 (cf. p. 58).
- [125] Richard NOWAKOWSKI et Peter M. WINKLER. « Vertex-to-vertex pursuit in a graph ». In : *Discr. Math.* 43 (1983), p. 235-239 (cf. p. 65, 90).
- [126] Damian OSAJDA et Piotr PRZYTYCKI. « Boundaries of systolic groups ». In : *Geometry & Topology* 13.5 (2009), p. 2807-2880 (cf. p. 66).
- [127] Megan OWEN et J. Scott PROVAN. « A Fast Algorithm for Computing Geodesic Distances in Tree Space ». In : *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8 (2011), p. 2-13 (cf. p. 63).
- [128] János PACH et Pankaj K AGARWAL. *Combinatorial Geometry*. T. 37. John Wiley & Sons, 2011 (cf. p. 10, 36).
- [129] Dömötör PÁLVÖLGYI et Gábor TARDOS. « Unlabeled compression schemes exceeding the VC-dimension ». In : *arXiv preprint arXiv :1811.12471* (2018) (cf. p. 40).
- [130] David PELEG. *Distributed Computing : A Locality-Sensitive Approach*. SIAM, 2000 (cf. p. 54).
- [131] David PELEG. « Proximity-preserving labeling schemes ». In : *J. Graph Theory* 33 (2000), p. 167-176 (cf. p. 11, 12, 51, 53, 54, 114).
- [132] David PELEG. « Informative labeling schemes for graphs ». In : *Theor. Comput. Sci.* 340 (2005), p. 577-593 (cf. p. 51, 114, 140).

-
- [133] David POLLARD. *Convergence of Stochastic Processes*. Springer Science & Business Media, 2012 (cf. p. [10](#), [44](#)).
- [134] Piotr PRZYTYCKI. « Systolic groups acting on complexes with no flats are word-hyperbolic ». In : *Fundamenta mathematicae* 193 (2007), p. 277-283 (cf. p. [66](#)).
- [135] Richard RADO. « Universal graphs and universal functions ». In : *Acta Arithmetica* 4.9 (1964), p. 331-340 (cf. p. [49](#)).
- [136] Neil ROBERTSON et Paul D SEYMOUR. « Graph minors. XX. Wagner's conjecture ». In : *J. Comb. Theory, Series B* 92.2 (2004), p. 325-357 (cf. p. [76](#)).
- [137] Benjamin I. P. RUBINSTEIN, Peter L. BARTLETT et J. Hyam RUBINSTEIN. « Shifting : One-inclusion mistake bounds and sample compression ». In : *J. Comput. Syst. Sci.* 75 (2009), p. 37-59 (cf. p. [40](#), [44](#)).
- [138] Gert SABIDUSSI. « Graph multiplication ». In : *Mathematische Zeitschrift* 72.1 (1959), p. 446-457 (cf. p. [28](#)).
- [139] Norbert SAUER. « On the density of families of sets ». In : *J. Comb. Theory, Series A* 13 (1972), p. 145-147 (cf. p. [10](#), [38](#), [42](#)).
- [140] Saharon SHELAH. « A combinatorial problem ; stability and order for models and theories in infinitary languages ». In : *Pac. J. Math.* 41.1 (1972), p. 247-261 (cf. p. [38](#)).
- [141] Sergey V. SHPECTOROV. « On scale embeddings of graphs into hypercubes ». In : *European J. Combin.* 14 (1993), p. 117-130 (cf. p. [32](#), [71](#)).
- [142] Daniel D SLEATOR et Robert Endre TARJAN. « A data structure for dynamic trees ». In : *J. Comput. Syst. Sci.* 26.3 (1983), p. 362-391 (cf. p. [50](#)).
- [143] Valeriu P. SOLTAN et Victor CHEPOI. « Conditions for invariance of set diameters under d-convexification in a graph ». In : *Cybernetics and Systems Analysis* 19.6 (1983), p. 750-756 (cf. p. [64](#)).
- [144] Mikkel THORUP. « Compact oracles for reachability and approximate distances in planar digraphs ». In : *J. of ACM* 51 (2004), p. 993-1024 (cf. p. [12](#)).
- [145] Mikkel THORUP et Uri ZWICK. « Compact routing schemes ». In : *SPAA*. ACM. 2001, p. 1-10 (cf. p. [54](#), [55](#)).
- [146] Mikkel THORUP et Uri ZWICK. « Approximate distance oracles ». In : *Journal of ACM* 52 (2005), p. 1-24 (cf. p. [12](#), [52](#)).
- [147] Leslie G VALIANT. « Universality considerations in VLSI circuits ». In : *IEEE Transactions on Computers* 100.2 (1981), p. 135-140 (cf. p. [49](#)).
- [148] Vladimir N. VAPNIK et Alexey Ya CHERVONENKIS. « On the uniform convergence of relative frequencies of events to their probabilities ». In : *Measures of complexity*. Springer, 2015, p. 11-30 (cf. p. [10](#), [33](#), [34](#), [36](#), [38](#)).

- [149] Vladim G. VIZING. « The Cartesian product of graphs (Russian) ». In : *Vycisl. Sistemy* 9 (1963), p. 30-43 (cf. p. 28).
- [150] Elke WILKEIT. « The retracts of Hamming graphs ». In : *Discr. Math.* 102.2 (1992), p. 197-218 (cf. p. 61).
- [151] Peter M. WINKLER. « Proof of the squashed cube conjecture ». In : *Combinatorica* 3 (1983), p. 135-139 (cf. p. 11, 51).
- [152] Peter M. WINKLER. « Isometric embedding in products of complete graphs ». In : *Discrete Appl. Math.* 7 (1984), p. 221-225 (cf. p. 26, 27).
- [153] Glynn WINSKEL. « An introduction to event structures ». In : *REX Workshop*. T. 354. Springer, 1988, p. 364-397 (cf. p. 61).

Index

- Symbols**
2-pulvérisation 98
2VC-dimension 98
- A**
ancêtre (arbre) 19
arête de type uv 24
arboricité 47
arbre 19
arbre des séparateurs récursifs 140
arbre étoilé 154
arête 18
arête de type uv 82
arête dominante 90
arête horizontale 95
arête verticale 95
augmentation d'une famille 95
- B**
bordure d'une fibre 124
boule 19
bouquet de cubes 31
bouquet de demi-cubes 32
branche (arbre) 19
- C**
c-pulvérisation (demi-cube) 99
carré d'un graphe 31
centroïde 53
chemin (graphe) 18
chenille (graphe) 19
clique plantée 96
clique-VC-dimension 101
clique-VC-dimension simple 99
clôture par mineur 76
coin (graphe planaire) 67
- complexe simplicial 31
complexe systolique 66
condition du quadrangle (CQ) . 57
condition du triangle (CT) 56
cône (graphe médian) 126
cône (graphe ponté) 156
convexe 20
coupe 25
cycle 18
- D**
d-décalage 102
décalage 41
dégénérescence 22
degré 18
degré moyen 21
degré moyen maximum 21
demi-cube 31
demi-espaces 27
densité 21
diamètre 19
différence symétrique 30
dimension 60
distance 19
- E**
empreinte 20
ensemble étoilé 154
ensemble/graphe k -porté 20
ensemble/graphe quasi-porté .. 20
espace CAT(0) 62
étoile (arbre) 19
étoile (graphe médian) 119
étoile (graphe ponté) 156
extension (produit cartésien) .. 74

- F**
- face (complexe simplicial) 31
 - face (graphe planaire) 67
 - famille d'ensembles ample 39
 - famille maximale 38
 - famille maximum 38
 - famille paire 95
 - famille plantée 96
 - feuille 19
 - fibre (demi-cube) 98
 - fibre (graphe médian) 122
 - fibre (produit cartésien) 74
 - fibre (graphe ponté) 156
 - fibres voisines 123
 - fil (arbre) 19
 - forêt (graphe) 19
 - frontière entre deux fibres 123
- G**
- graphe (simple) 18
 - graphe avec ports 54
 - graphe biparti 18
 - graphe complet (ou clique) 18
 - graphe connexe 19
 - graphe cordal 64
 - graphe de 1-inclusion 31
 - graphe de 1,2-inclusion 95
 - graphe de Hamming 25
 - graphe de Johnson 31
 - graphe démontable 90
 - graphe faiblement modulaire .. 56
 - graphe fini 18
 - graphe \mathcal{G} -universel 49
 - graphe \mathcal{G} -universel induit 49
 - graphe irréductible 29
 - graphe k -parti 18
 - graphe k -parti complet 18
 - graphe médian 58
 - graphe modulaire 59
 - graphe non-orienté 18
 - graphe ponté 64
 - graphe premier 28
 - graphe quasi-médian 61
- graphe quotient 29
- graphe sans boucle 18
- graphe sans cube 116
- graphe simplexe 119
- graphe trivial 18
- graphes isomorphes 23
- H**
- hauteur (arbre) 19
 - homomorphisme de graphes ... 23
 - hypercube 25, 30
- I**
- intervalle 19
 - involution 91
 - isomorphisme de graphes 23
- K**
- k -dégénéré 22
 - k -ensemble 30
 - k -roue 19
 - k -simplexe 31
- L**
- localement convexe 59
- M**
- médian d'un triplet 58
 - mineur de graphe 76
 - modèle RAM 46
- O**
- octaèdre 91
- P**
- panneau (graphe médian) 126
 - panneau (graphe ponté) 156
 - parent (arbre) 19
 - plongement avec distorsion 23
 - plongement irredondant 29
 - plongement isométrique 23
 - port 54
 - porte 20
 - profondeur (arbre) 19
 - projection (demi-cube) 98

-
- projection (produit cartésien) .. 74
 projection métrique.....19
 pulvérisation (produit cartésien)
 76
 pulvérisation 34
 pulvérisation de mineur (produit
 cartésien) 77
- Q**
- quasi-médian.....58
 quasi-représentant 168
- R**
- racine 19
 relation Θ 27
 représentant 168
 rétracté 23
 rétraction 23
- S**
- s-pulvérisation 108
 schéma de distance 51
 schéma d'adjacence.....46
 schéma d'étiquetage..... 46
 schéma de distance k -approché
 148
 schéma de routage 54
 sommet 18
 sommet de type u 24
 sommet dominé 90
 sommet feuille 83
 sommet interne (graphe planaire)
 67
 sommet médian 53
 sommet simplicial 64
- sommets 1-voisins..... 130
 sommets 1pc-voisins.....167
 sommets 2-voisins..... 130
 sommets adjacents/voisins.....18
 sommets éloignés 130
 sommets proches..... 130
 sous-cube 74
 sous-graphe 18
 sous-graphe induit.....20
 sous-graphe isométrique..... 20
 sous-graphe porté 20
 sous-produit.....74
 sous-produit mineur..... 76
 star-VC-dimension.....109
 star-VC-dimension simple.....109
- T**
- trace (produit cartésien) 74
 trace (ensembles) 34
 triangle métrique 57
 triangle métrique équilatéral...57
 triangle plat.....150
- U**
- (u, v) -chemin 18
- V**
- VC-densité 77
 VC-densité mineure.....77
 VC-dimension mineure 77
 VC-dimension 34
 VC-dimension induite (produit
 cartésien) 77
 voisinage fermé.....21
 voisinage ouvert 21

Table des figures

1.1	Graphes et notations usuels	19
1.2	Notions de sous-graphes.	21
1.3	Pour la proposition 1.	21
1.4	Isomorphisme et rétraction.	23
1.5	Produit cartésien et arête de type uv	24
1.6	Plongement isométrique dans l'hypercube.	28
1.7	Factorisation en produit de facteurs premiers.	29
1.8	Plongement isométrique canonique.	30
2.1	Fibres et extensions.	37
2.2	Définition de VC-dimension par contraction.	37
2.3	Plongement et VC-dimension.	38
3.1	Étiquetage d'adjacence d'un graphe d'arboricité 2.	48
3.2	Graphe universel induit.	50
3.3	Schéma de distance pour les arbres.	54
4.1	Conditions du triangle et du quadrangle.	57
4.2	Graphe médian et graphe ponté.	60
4.3	Axiome de comparaison Cartan-Alexandrov-Toponogov.	62
4.4	Condition du 3-cube	63
4.5	Convexité des boules dans les graphes pontés.	65
4.6	Illustration du théorème de Gauss-Bonnet.	68
5.1	Sous-produit et fibre.	75
5.2	Sous-produits mineurs et pulvérisation	78
5.3	Exemple d'un graphe G tel que $VC\text{-dens}^*(G) > VC\text{-dens}(G)$	79
5.4	G, G_{uv}, G^{uv} et G_c^{uv}	83
5.5	VC-dimension induite et contraction	86
5.6	Preuve du lemme 14	87
5.7	Preuve du lemme 16	93
6.1	Opération de plantation d'une clique.	97
6.2	Une paire c -pulvérisée.	99
6.3	Illustration de l'exemple 5.	101
6.4	Preuve du lemme 21.	105

7.1	Deux exemples de graphes médians	117
7.2	Un graphe et son graphe simplexe	120
7.3	Étiquetage de l'étoile d'un sommet.	121
7.4	Preuve des lemmes 25 et 26	122
7.5	Preuve du lemme 32.	129
7.6	Classification des paires de sommets.	130
7.7	Plus courts chemins.	132
7.8	Partie i de $LD(u)$	135
8.1	Pour le lemme 38.	152
8.2	Grille triangulaire rongée	153
8.3	Étoile dans un graphe ponté sans K_4	156
8.4	Illustration du lemme 44.	158
8.5	Les fibres des graphes pontés sans K_4 sont isométriques.	160
8.6	Aide à la lecture des quatre cas pour montrer $t \in \Pi$ dans la preuve du lemme 49.	164
8.7	Pour le lemme 51.	166
8.8	Pour la preuve du lemme 52.	168
8.9	Corollaire 11 et lemme 54.	169
8.10	Pour la preuve de l'affirmation 54.1.	172
8.11	Encodage d'une étoile dans un graphe ponté sans K_4	173