Problèmes d'identification métrique dans les graphes

Metric identification problems in graphs

David Auger

2007D013

Octobre 2007

Département Informatique et Réseaux
Groupe MIC2 : Mathématiques de l'Information,
des Communications et du Calcul

Problèmes d'identification métrique dans les graphes

David AUGER
École nationale supérieure des télécommunications
46 rue Barrault, 75634 Paris Cedex 13, France
auger@enst.fr

Résumé

Nous étudions quelques problèmes de *discrimination* dans les graphes. Après avoir présenté le cadre général, nous nous intéressons plus spécifiquement aux *codes identifiants*, et en particulier à la complexité algorithmique du calcul d'un code identifiant de cardinal minimum. Nous abordons ensuite l'étude de quelques propriétés structurelles des *graphes sans jumeaux*, c'est-à-dire admettant un code identifiant. En particulier, nous prouvons une conjecture selon laquelle tout graphe sans r-jumeaux admet une chaîne à $2r+1$ sommets en tant que sous-graphe. Enfin, dans une dernière partie nous évoquons deux variantes des codes identifiants, les *codes localisateurs-dominateurs* et les *codes de position*.

Metric Identification Problems in Graphs

David AUGER
Ecole nationale supérieure des télécommunications
46 rue Barrault, 75634 Paris Cedex 13, France
auger@enst.fr

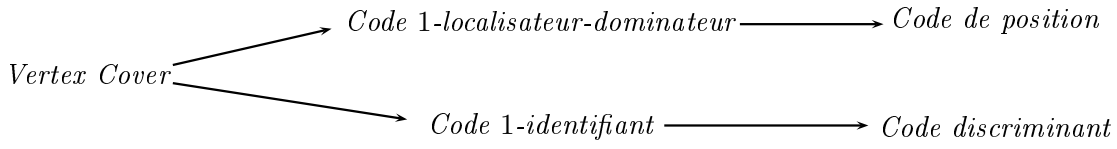
Résumé

We study some *discrimination* problems in graphs. After an introduction to the general framework, we focus on *identifying codes* and on the algorithmic complexity of searching such a code of minimal size. We then study a few structural properties of *twin-free graphs*, wich are precisely those graphs in which identifying codes exist. We prove a conjecture which asserts that a r-twin-free graph admits a path on $2r+1$ vertices as an induced subgraph. The last part is devoted to variants of identifying codes such as *locating-dominating codes* and *position codes*.

Ce travail a été réalisé pour un mémoire de stage du Master 2 Modélisation et Méthodes Mathématiques en Economie et Finance de l'Université Paris I Panthéon - Sorbonne, en spécialité Recherche Opérationnelle. Ce stage a été effectué de mai à septembre 2007 à l'Ecole Nationale Supérieure des Télécommunications de Paris, sous la direction d'Olivier Hudry. Nous avons étudié le problème de la recherche de codes r-identifiants dans les graphes ainsi que les paramètres structurels des graphes sans r-jumeaux, c'est-à-dire admettant un code r-identifiant.

Dans une première partie, nous allons présenter la classe plus large des problèmes de discrimination, liés aux codes discriminants dans les graphes bipartis ; la deuxième partie sera ensuite plus particulièrement consacrée aux codes identifiants. Dans une troisième partie, nous présenterons quelques résultats concernant les paramètres structurels des graphes sans r-jumeaux ; enfin, une dernière partie présentera deux autres types de codes métriques pour les graphes.

Il s'avère que la recherche des codes de cardinal minimum, pour chacun des quatre types de codes présentés, constitue un problème NP-difficile. Nous effectuons les preuves de ces faits par réduction polynomiale, en partant du problème Vertex Cover - qui fait partie des vingt-et-un problèmes dont la NP-complétude a été prouvée par R. Karp en 1972 - en procédant selon l'ordre suivant :



Toutes les preuves qui figurent dans ce mémoire sont (plus ou moins) originales, mais reprennent des résultats déja connus ; la seule exception étant le théorème 18 - dont la preuve figure en annexe pour ne pas alourdir le texte - qui établit la validité d'une conjecture publiée en 2006 par Charon, Honkala, Hudry et Lobstein.

Table des matières

1 Quelques problèmes de discrimination

1.1 "Qui est-ce ?"

Le principe du jeu "Qui est-ce ?", distribué en France par MB dans les années 1980, est le suivant : deux joueurs s'affrontent en essayant de deviner l'individu qui figure sur la carte de son adversaire, choisie au hasard parmi vingt-quatre. A tour de rôle, chacun d'eux doit poser à son adversaire une question relative aux attributs physiques des individus, comme par exemple : "A-t-il une casquette ?", "Est-ce une fille ?" ou encore "A-t-il a une moustache ?".

FIG. 1 – Max, un des personnages du "Qui est-ce ?"

Pour gagner, un joueur doit tenter de minimiser de façon efficace le nombre de questions qu'il va poser. Nous allons ici envisager une variante simplifiée de ce jeu et nous intéresser à la minimisation du nombre de questions.

Dans notre jeu ne figurent que cinq *individus* qui sont Aude, Benoît, Cécile, Didier ou Elise. Pour les identifier nous pourrons poser des questions relatives à cinq *attributs*, qui sont :

- être de sexe féminin (F) ;
- avoir les yeux bleus (B) ;
- avoir les cheveux courts (C) ;
- avoir des tâches de rousseur (R) ;
- porter des lunettes (L).

Chacun des individus possède ou non chaque attribut, conformément au tableau suivant :

	(F) : Féminin	(B) : Y. bleus	(C) : Ch. courts	(R) : T. Rousseur	(L) : Lunettes
Aude	•		•		•
Benoît			•		•
Cécile	•	•	•		
Didier					•
Elise	•		•	•	

Dans le jeu de MB, les joueurs établissent leurs questions au fur et à mesure des réponses qu'ils vont obtenir. Pour le problème qui va nous intéresser, nous allons au contraire nous placer dans une situation *statique* : nous devons établir les questions avant le début du jeu et obtiendrons toutes les réponses d'un coup. Précisons un peu les règles :

(i) Nous choisissons entre une et cinq questions parmi cinq questions possibles qui sont du type "la personne choisie possède-t-elle l'attribut X ?". Nous composons ainsi un questionnaire.

(ii) L'adversaire répond au questionnaire. Il répond à toutes nos questions et il lui est interdit de mentir.

(iii) Etant donné les réponses de l'adversaire nous tentons de deviner quel individu il cache. Celui qui a deviné juste en ayant composé le questionnaire comportant le moins de questions gagne la partie.

Dans cet exemple simple, on voit que pour gagner à coup sûr (ou au pire être ex-aequo) il suffit de composer un questionnaire de taille minimum parmi les questionnaires identifiant sans doute la carte de l'adversaire. Le problème est donc le suivant :

Comment choisir un ensemble d'attributs permettant d'identifier chaque individu, et dont le cardinal est le plus petit possible ?

Pour commencer, remarquons que cinq questions permettent toujours d'identifier les individus, car il n'en existe pas deux qui présentent exactement les mêmes attributs ; il semble cependant que l'on puisse faire mieux, en éliminant par exemple l'attribut (L). Le tableau résultant de cette élimination est le suivant :

	(F) : Féminin	(B) : Y. bleus	(C) : Cheveux courts	(R) : Tâches
Aude	•		•	
Benoît			•	
Cécile	•	•	•	
Didier				
Elise	•		•	•

Comme on le voit, ces quatre attributs suffisent encore à identifier les individus ; soit quatre questions pour le questionnaire. Peut-on faire mieux en éliminant à la suite un autre attribut ? Non, car :

- (F) est maintenant le seul attribut permettant de distinguer Aude de Benoît, il est donc nécessaire de conserver la question correspondante ;
- de même (B) est le seul attribut permettant de distinguer Aude de Cécile ;
- (C) est maintenant le seul qui permette de distinguer Benoît de Didier ;
- enfin (R) est le seul attribut permettant de distinguer Aude d'Elise.

Nous avons donc atteint une configuration *minimale* de quatre attributs, où aucun attribut ne peut plus être supprimé ; il pourrait sembler logique de conclure que le nombre minimum de questions à poser est quatre.

Pourtant, on peut faire mieux en commençant autrement : l'élimination au départ des attributs (F) et (R) aurait conduit au tableau suivant :

	(B) : Y. bleus	(C) : Cheveux courts	(L) : Propriétaire
Aude		•	•
Benoît	•	•	•
Cécile	•	•	
Didier			•
Elise		•	

qui nous permet bien d'identifier chaque individu par ses attributs. Nous parvenons donc à trois questions, et pouvons aisément conclure que ceci est le minimum possible, puisqu'on ne peut répondre à deux questions que de quatre manières différentes (Oui/Oui, Oui/Non, Non/Oui, Non/Non), et ainsi deux questions ne peuvent permettre d'identifier cinq individus : il en faut au moins trois, d'où l'optimalité de la solution ci-dessus.

Nous appellerons *problèmes de discrimination* les problèmes de ce type. Il s'agira toujours de choisir un ensemble d'attributs le plus petit possible de manière à pouvoir identifier des individus ; les individus et attributs pouvant être de nature quelconque. L'exemple précédent pouvant sembler quelque peu arbitraire, nous présentons au paragraphe suivant une application possible à la bioinformatique.

1.2 Diagnostic automatique d'une maladie

Considérons un dictionnaire médical, présentant des maladies (les individus) et pour chacune d'elles un ensemble de symptômes (les attributs) lui correspondant. Nous voulons pouvoir, étant donné les symptômes présentés par un patient, identifier sans doute possible la maladie qu'il a contracté (en supposant bien sûr que celle-ci soit recensée par notre dictionnaire).

Une telle situation pourrait se présenter, par exemple, pour la réalisation d'un outil de diagnostic automatique : le praticien n'aurait qu'à constater de manière systématique la présence ou non de symptômes d'une liste donnée sur le patient, et le dictionnaire identifierait la maladie. Toutefois, la détection de certains symptômes pouvant nécessiter du matériel, il est irréaliste de demander au praticien d'être prêt à vérifier la présence d'un grand nombre de symptômes : nous avons donc interêt à choisir le plus petit ensemble de symptômes possible, qui permette d'identifier à coup sûr la maladie. Il s'agit donc un problème de discrimination au sens du paragraphe précédent.

1.3 Discrimination dans un ensemble abstrait

Formalisons de manière plus générale les problèmes de discrimination que nous venons d'évoquer.

Nous nous donnons deux ensembles finis I et A appelés respectivement *ensemble des individus* et *ensemble des attributs*. A chaque $a \in A$ correspond une partie F_a de I, représentant l'ensemble des individus possédant l'attribut a. Pour tout individu $x \in I$, nous noterons $A(x)$ l'ensemble des attributs de x :

$$A(x) = \{a \in A \text{ t.q. } x \in F_a\}.$$

La famille $\mathcal{F} = \{F_a\}_{a \in A}$ sera dite *discriminante* si

$$\forall x, y \in X,\ x \neq y \Longrightarrow A(x) \neq A(y).$$

Le problème général alors peut être formulé de la façon suivante :

Etant donné une famille discriminante $\mathcal{F}$ sur un ensemble fini I, déterminer une famille discriminante $\mathcal{F}' \subset \mathcal{F}$ qui soit de cardinal minimum. (1)

Nous verrons que ce problème dont l'énoncé est assez simple est pourtant difficile à résoudre en toute généralité.

1.4 Codes discriminants

Le problème des *codes discrimants* est une reformulation du problème fondamental (1) avec le vocabulaire des graphes bipartis (voir la section 2.2 pour un rappel des définitions concernant les graphes). En conservant les notations précédentes, on considère le graphe biparti dont l'ensemble de sommets est $I \cup A$; l'ensemble des arêtes est constitué des paires (i, a) où l'individu i possède l'attribut a. En reprenant notre premier exemple, nous obtenons le graphe de la figure 2.

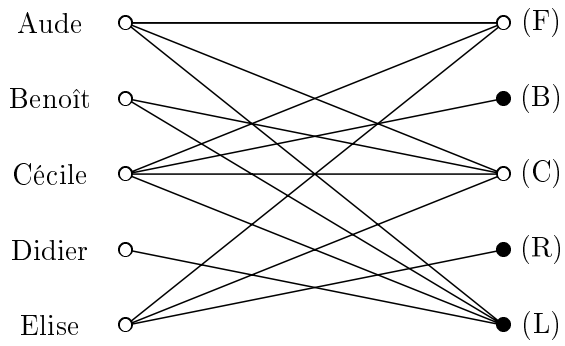


FIG. 2 – Un exemple de code discriminant - les mots de code sont en noir.

Pour les problèmes de discrimination dans les graphes, nous utilisons un vocabulaire issu de la théorie du codage : les attributs choisis correspondent à des sommets qui seront appelés des *mots de codes* ou plus simplement *codes*. On dira qu'un mot de code c couvre un sommet x si x possède l'attribut c; enfin, un mot de code c *sépare* deux sommets x et y si c couvre un et un seul des deux sommets x et y. Intuitivement, un mot de code séparant deux sommets est un attribut permettant de différentier ces deux sommets.

Plus formellement, un *code discriminant* sur un graphe biparti $G = (I \cup A, E)$ est une partie $\mathcal{C}$ de A telle que pour tous $x, y \in I$ on ait

$$x \neq y \Longrightarrow B(x,1) \cap \mathcal{C} \neq B(y,1) \cap \mathcal{C}$$

où $B(x,1)$ désigne la boule de centre x et de rayon 1, c'est-à-dire l'ensemble de sommets constitué des voisins de x et de x lui-même. D'après notre étude, le code constitué par les sommets (B), (R) et (L) (en noir sur la figure 2) est un code discriminant dont le cardinal est minimum. Le problème de discrimination (1) peut alors se reformuler sous la forme suivante :

Etant donné un graphe biparti $G = (I \cup A, E)$, déterminer une partie $\mathcal{C} \subset A$ qui soit un code discriminant dont le cardinal est minimum. (2)

Nous avons le résultat élémentaire suivant :

Proposition 1. *Soit $G = (I \cup A, E)$ un graphe biparti et $\mathcal{C} \subset A$ un code discriminant. Alors*

$$|\mathcal{C}| \geq \lceil \log_2(|I|) \rceil$$

et cette inégalité est optimale.

Preuve. Notons $K(x)$ pour tout $x \in I$ l'ensemble $B(x,1) \cap \mathcal{C}$. Par définition d'un code discriminant l'application

$$K : I \longrightarrow 2^{\mathcal{C}}$$

est injective et donc $2^{|\mathcal{C}|} \geq |I|$, d'où le résultat en considérant le fait que $|\mathcal{C}|$ est entier. D'autre part, il est clair qu'en prenant un ensemble $\mathcal{C}$ de codes pour attributs et $I = 2^{\mathcal{C}}$, l'inégalité est atteinte.

□

1.5 Complexité du problème

Rappelons qu'un problème de décision, c'est-à-dire admettant une réponse de type Oui/Non, est dit appartenir à la classe P (pour *Polynomial*) s'il existe un algorithme permettant de le

résoudre en un temps d'exécution majoré dans le pire des cas par un polynôme en la taille des données. Ces problèmes sont considérés comme faciles à résoudre au sens la durée d'exécution de l'algorithme augmente à vitesse raisonnable lorsque la taille de l'instance augmente. Un problème qui n'est pas dans la classe P est dit *exponentiel*. Un problème de décision est dit dans la classe NP (pour *Non-deterministic Polynomial*) s'il est toujours possible de certifier que la réponse est "oui" en donnant un certificat dont l'exactitude peut être vérifiée en temps polynomial. Par exemple, il est très difficile de déterminer s'il est possible de colorier avec trois couleurs seulement les sommets d'un (grand) graphe de sorte que deux sommets adjacents ne soient jamais de la même couleur ; mais si la réponse est "oui"', on peut le prouver en exhibant un tel coloriage, et il est alors rapide de vérifier que ce coloriage satisfait bien aux conditions exigées.

Enfin, un problème de décision est dit $NP-complet$ s'il appartient à la classe NP et s'il est, en un certain sens, plus dur que tous les problèmes de la classe NP. Ces problèmes sont considérés très difficiles au sens où il apparaît à l'heure actuelle peu probable de trouver un jour un algorithme polynomial pour les résoudre. Pour plus de détails et la formalisation des notions précédentes, nous renvoyons à [8].

Le problème de décision associé à la recherche d'un code discriminant de cardinal minimum est le suivant :

DISCRIMINATION
INSTANCE : Un graphe biparti $G = (A \cup I, E)$; un entier $k \geq 1$.
QUESTION : G admet-t-il un code discriminant $\mathcal{C} \subset A$ tel que $|\mathcal{C}| \leq k$?

Le théorème qui suit s'applique alors problème ci-dessus ou au problème sous forme ensembliste (1).

Théorème 2. *Le problème* DISCRIMINATION *est NP-complet.*

Nous prouverons ce résultat à la suite du théorème 11 (page 24).

2 Codes identifiants

2.1 Identification des pannes dans un réseau de processeurs

Considérons un réseau de processeurs disposés selon un circuit éléctrique planaire, conformément à la figure 3. Certains processeurs sont reliés sur ce schéma par une arête, indiquant une proximité électronique dans le circuit ; nous dirons qu'ils sont *voisins*. Dans ce réseau, les processeurs peuvent tomber en panne et notre but est de les équiper de façon efficace de *récepteurs* de façon à localiser le processeur défectueux. Nous supposerons qu'il ne peut y avoir qu'au plus *une seule panne à la fois*.

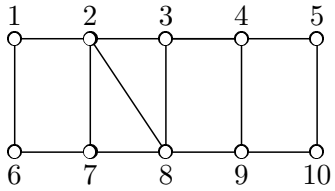


FIG. 3 – Un réseau de processeurs

Imaginons que nous placions un récepteur sur le processeur 2. La figure 4 représente la *zone de couverture* de ce récepteur, composée du processeur 2 lui-même et de ses voisins. Si une panne se produit dans la zone de couverture d'un récepteur, celui-ci nous envoie un signal électrique. L'information dont nous disposerions dans le cas d'un récepteur placé en 2 serait alors précisément la suivante :

Il y a une panne dans le réseau, et le processeur défectueux est le 1*, le* 2*, le* 3*, le* 7 *ou le* 8*.*

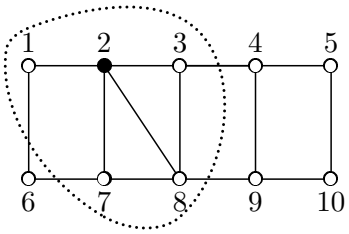


FIG. 4 – Un récepteur placé en 2 couvre les processeurs 1, 2, 3, 7 et 8.

Résumons les caractéristiques des récepteurs :

(i) ils peuvent être placés sur chacun des processeurs ;

(ii) ils sont capables de détecter si le processeur sur lequel ils sont placés, ou un des processeurs voisins, est défectueux, de manière infaillible ;

(iii) le cas échéant, ils nous envoient un signal nous indiquant la détection d'une panne ;

(iv) le signal envoyé est de type binaire (impulsion électrique) et n'indique pas quel récepteur est en panne, mais simplement que le recepteur a détecté une panne dans sa zone de couverture.

De manière à localiser les pannes, il nous faut placer dans le réseau plusieurs récepteurs pour pouvoir recouper les informations. Par exemple, placer un récepteur sur les processeurs 2 et 4 permettrait de savoir si le processeur 3 est en panne (voir fig. 5), puisqu'il est le seul à être couvert par ces deux récepteurs. Mais est-il possible de pouvoir identifier tous les processeurs ?

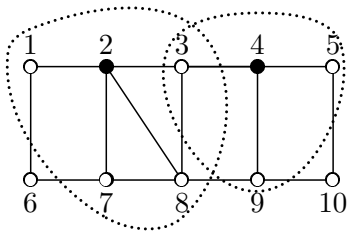


FIG. 5 – Deux récepteurs placés en 2 et en 4 permettent d'identifier le récepteur 3

On peut remarquer que la réponse à cette question est positive dès lors qu'on place un récepteur sur chaque processeur. En effet, il suffit de dresser la liste, pour chaque processeur x, des processeurs qui seront couverts par un détecteur placé en x :

	1	2	3	4	5	6	7	8	9	10
1	•	•				•				
2	•	•	•				•	•		
3		•	•	•				•		
4			•	•	•				•	
5				•	•					•
6	•					•	•			
7		•				•	•	•		
8		•	•				•	•	•	
9				•				•	•	•
10					•				•	•

Puisqu'il n'existe pas, comme on peut le voir ci-dessus, deux récepteurs couverts par exactement les mêmes détecteurs, l'identification du processeur défectueux sera possible. Remarquons aussi que tous les processeurs sont couverts par au moins un détecteur, de sorte qu'une panne existera dans le réseau si et seulement si un signal a été reçu. Prenons quelques exemples :

- Les récepteurs placés en 1, 6 et 7 envoient un signal, indiquant la présence d'une panne dans leurs zones de couverture. D'après notre tableau c'est le processeur 6 qui présente une panne.
- De même, si les récepteurs placés en 2, 3, 4 et 8 envoient un signal, c'est que c'est le processeur 3 qui est en panne.
- S'il n'y a pas de signal, c'est qu'il n'y a pas de panne.
- Si nos récepteurs fonctionnent correctement et qu'il n'y a qu'une seule panne, il est impossible que nous recevions le signal 1, 2, 4, 9 puisqu'il ne correspondrait à aucun récepteur.

A partir de ce premier constat, et dans l'hypothèse que chaque chaque récepteur a un coût non négligeable, nous pouvons tenter de minimiser le nombre de récepteurs. Un examen du tableau ci-dessus montre que nous avons affaire à un problème de discrimination au sens de la section 1 ; les individus sont les sommets et les attributs correspondent aux propriétés "être couvert par le sommet x". Il subsiste une légère différence qui est qu'on exige ici que chaque processeur soit couvert par au moins un récepteur, alors que pour les problèmes de discrimination généraux un individu (au plus) pouvait très bien ne posséder aucun des attributs sélectionnés.

Pour l'exemple ci-dessus, il est déja assez difficile de trouver un ensemble de processeurs qui soit identifiant et de cardinal minimum. Par exemple, on peut se rendre compte que le choix des

sommets 6, 7, 8, 9 et 10 conviendrait :

	6	7	8	9	10
1	•				
2		•	•		
3			•		
4				•	
5					•
6	•	•			
7	•	•	•		
8		•	•	•	
9			•	•	•
10				•	•

Cette configuration de récepteurs est *minimale*, dans le sens où aucun récepteur ne peut plus être supprimé. Par exemple, en supprimant le récepteur 10 on ne pourrait plus distinguer les récepteurs 4 et 10 l'un de l'autre.

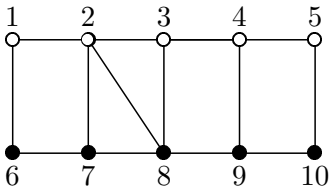


FIG. 6 – Un choix minimal pour les récepteurs, mais de cardinal pas forcément minimum.

Cette configuration est-elle pour autant *minimum* ? Un argument simple, que nous verrons dans la prochaine section, montre qu'il faut au moins 4 récepteurs pour identifier un réseau de 10 processeurs. Sur cet exemple simple, nous sommes déja relativement démunis ; on pourrait par exemple tester toutes les configurations à 4 récepteurs (soit 210 possibilités), pour être sûrs qu'on ne peut faire mieux que notre configuration à 5 récepteurs. Nous entrevoyons ici la forte complexité algorithmique de ce problème, qui sera étudiée au 2.8.

2.2 Rappels sur les graphes

Graphes, sommets et arêtes.

Nous considérons dans ce mémoire uniquement des *graphes simples non-orientés* , ce que nous appelerons plus simplement un *graphe*. Formellement, un graphe G est donc un couple (V, E) où V est un ensemble (non-vide) dont les éléments sont appelés *sommets* du graphe et E un ensemble de parties à deux éléments de V, appelées *arêtes* du graphe. Une arête $\{x, y\}$ avec x et y dans V est plus simplement notée xy. Une *chaîne* dans le graphe G est une suite finie de sommets $v_0, v_1, \cdots v_k$ tels que $v_i v_{i+1} \in E$ pour tout $0 \leq i \leq k-1$; on notera $v_0 v_1 \cdots v_k$ une telle chaîne. Les sommets v_0 et v_k sont les *extrémités* de cette chaîne, et sa longueur est k (nombre d'arêtes).

On dit que G est connexe si pour tout couple de sommets distincts (x, y), il existe une chaîne dont les extrémités sont x et y. La plupart des graphes que nous allons considérer seront connexes.

Un graphe $G = (V, E)$ est dit *biparti* s'il est possible de partitionner les sommets en deux ensembles non-vides A et B, de sorte qu'une arête de G ait toujours une extrémité dans A et une extrémité dans B.

Chaînes, distances, excentricité, rayon et diamètre.

On munit V d'une distance naturelle en définissant pour tous $x, y \in V$ la distance $d(x, y)$ comme étant la plus petite longueur possible d'une chaîne $v_0 v_1 \cdots v_k$ *joignant* x à y, c'est-à-dire telle que $v_0 = x$ et $v_k = y$. Autrement dit, deux sommets x et y sont à distance k s'il est possible de passer de x à y en suivant k arêtes consécutives, mais pas moins. Si G n'est pas connexe, on peut poser $d(x, y) = \infty$ dans le cas où x et y ne peuvent être joints par une chaîne.

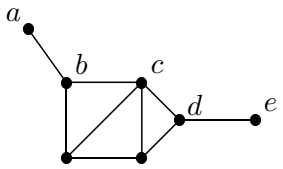


FIG. 7 – Un graphe G. Les sommets a et e sont à distance 4 car $abcde$ est une plus courte chaîne joignant a et e ; le diamètre de G est aussi 4. Le rayon vaut 2, l'unique centre étant c.

L'*excentricité* d'un sommet $x \in V$ est la plus grande distance entre x et les autres sommets du graphe

$$\mathrm{exc}(x) = \max_{v \in V} d(x, v).$$

Les sommets de G ayant une excentricité minimum sont appelés *centres* du graphe ; la valeur de leur excentricité est le *rayon* de G, noté $\mathrm{rad}(G)$. De même, l'excentricité maximale dans le graphe est le *diamètre* $\mathrm{diam}(G)$ de G, ce qui peut aussi s'écrire

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

Sous-graphes et graphes partiels.

Un *graphe partiel* (en anglais *subgraph*) de $G = (V, E)$ est un graphe $H = (W, F)$ tel que $W \subset V$ et $F \subset E$. Un *sous-graphe* (en anglais *induced subgraph*) de G est un graphe $H = (W, F)$ tel que $W \subset V$ et

$$F = \left\{ xy,\ (x, y) \in W^2 \text{ et } xy \in E \right\}.$$

Autrement dit, on obtient un sous-graphe de G en éliminant des sommets mais en conservant toutes les arêtes entre les sommets restants ; alors qu'un graphe partiel est obtenu en éliminant des sommets ainsi que des arêtes (voir l'exemple en figure 8).

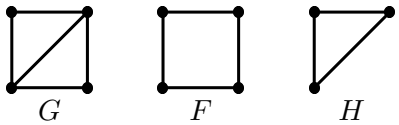


FIG. 8 – F est un graphe partiel de G mais pas un sous-graphe ; H est un sous-graphe de G

Etant donnée une chaîne $P = a_1 \cdots a_k$ où les a_i sont de sommets de G, dire que P est *sans corde* signifie que P est un sous-graphe de G.

2.3 Codes identifiants

Les codes identifiants ont été introduits pour la première fois en 1998 par Karpovsky, Chakrabarty et Levitin dans [11]. Il s'agit d'un problème de discrimination des sommets d'un graphe à l'aide des distances séparant un sommet des mots de code, correspondant à l'exemple des réseaux de processeurs que nous évoquions en introduction.

Soit $G = (V, E)$ un graphe. Nous dirons qu'un sommet c *couvre* un sommet x si $d(c, x) \leq 1$, et plus généralement qu'il r-couvre x si $d(x, c) \leq r$, ceci pour tout $r \geq 1$. L'ensemble des sommets qui sont r-couverts par c est la *boule* de centre c et de rayon r, notée $B(c, r)$. Nous dirons que deux sommets x et y sont *séparés* par c si c couvre un et un seul des deux sommets x et y. Un *code séparant* de G est une partie $\mathcal{C}$ de V telle que pour tous x et y, il existe $c \in C$ séparant x et y. Les éléments de c seront appelés *mots de code* ou plus simplements *codes*.

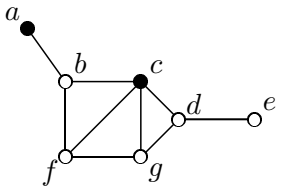


FIG. 9 – Le sommet c couvre les sommets b, c, d, f et g. Le sommet a 2-couvre les sommets a, b, c et f ; il 2-sépare (en particulier) les couples de sommets (c, d), (b, g) ou encore (a, e).

Une autre de manière de voir les choses est de remarquer qu'une partie $\mathcal{C}$ de V est un code séparant si et seulement si les n ensembles de sommets

$$B(x, 1) \cap \mathcal{C}$$

pour $x \in V$ sont distincts (où $n = |V|$). De la même manière on définit pour $r \geq 1$ un code *r-séparant* comme une partie $\mathcal{C} \subset V$ telle que les n ensembles de sommets

$$B(x, r) \cap \mathcal{C}$$

pour $x \in V$ soient distincts, comme on peut aussi dire que pour tout choix de deux sommets distincts x et y, il existe un mot de code $c \in \mathcal{C}$ qui r-couvre x et ne r-couvre pas y, ou l'inverse.

Un code $\mathcal{C} \subset V$ est dit *couvrant* (respectivement *r-couvrant*) si pour tout $v \in V$ il existe $c \in \mathcal{C}$ tel que $d(v, c) \leq 1$ (respectivement $d(v, c) \leq r$). Il est équivalent de demander que

$$\bigcup_{c \in \mathcal{C}} B(c, r) = V.$$

La définition classique des codes identifiants est la suivante :

Définition 3. *Un code r-identifiant est un code à la fois r-séparant et r-couvrant.*

De manière équivalente, un code est r-identifiant si les pour tous $v \in V$ les ensembles

$$B(v, r) \cap \mathcal{C}$$

sont non vides et distincts. En pratique, il peut être commode de confondre les codes identifiants et séparants sans se soucier de la couverture. En effet, si $\mathcal{C}$ est un code r-identifiant alors il est r-séparant ; réciproquement si $\mathcal{C}$ est un code r-séparant alors il existe au plus un sommet c tel que $B(c, r) \cap \mathcal{C}$ soit vide : alors $\mathcal{C} \cup \{c\}$ est un code r-identifiant.

2.4 Graphes sans jumeaux

Tous les graphes admettent un code couvrant, mais tous n'admettent pas de code séparant (ou identifiant). Quand il en existe un, nous dirons que le graphe est *identifiable* (ou *r-identifiable*). Un exemple simple de graphe non identifiable est le graphe complet K_n, constitué de n sommets tous reliés entre eux par une arête (voir figure 10). En effet quel que soient $x \in V$ et $r \geq 1$, la boule $B(x, r)$ est l'ensemble V tout entier donc quel que soit nos choix tous les sommets seront couverts par tous les codes, ce qui ne nous permettra aucune identification.

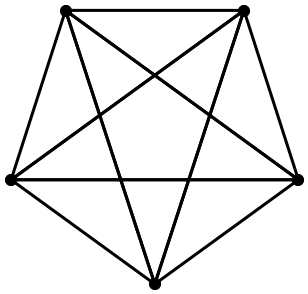


FIG. 10 – Le graphe complet K_5

Sans aller aussi loin, on peut se rendre aisément compte que le graphe de la figure 11 n'est pas identifiable : les sommets notés a et b ont même boule de rayon 1, et aucun code ne permettra de les distinguer. Deux tels sommets seront dits *jumeaux*.

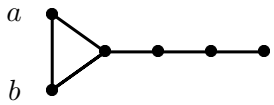


FIG. 11 – Deux jumeaux dans un graphe : $B(a, 1) = B(b, 1)$.

Un *graphe sans jumeaux* (respectivement, sans r-jumeaux) est donc un graphe $G = (V, E)$ tel que

$$\text{pour tous } x, y \in V, \ \ x \neq y \Longrightarrow B(x, 1) \neq B(y, 1)$$

(respectivement $B(x, r) \neq B(y, r)$). Nous pouvons maintenant caractériser les graphes r-identifiables :

Théorème 4. *Soit $r \geq 1$. Un graphe est r-identifiable si et seulement si il est sans r-jumeaux.*

Preuve. Si un graphe est sans r-jumeaux, alors les boules $B(x, r)$ pour $x \in V$ sont toutes distinctes et dont V tout entier est un code r-identifiant. Réciproquement, s'il existe x distinct de y tel que $B(x, r) = B(y, r)$, alors on aura $B(x, r) \cap \mathcal{C} = B(y, r) \cap \mathcal{C}$ quel que soit le choix de $\mathcal{C}$; G ne peut donc admettre de code r-séparant, et a fortiori de code r-identifiant. □

Etant donné un graphe, on peut se demander pour quelles valeurs de $r \geq 1$ il est r-identifiable. La proposition suivante aide à répondre à cette question :

Proposition 5. *Soit $G = (V, E)$ un graphe, $r \geq 1$ un entier et $a, b \in V$. Si a et b sont r-jumeaux alors ils sont $(r + 1)$-jumeaux.*

Corollaire 6. *Si G est r-identifiable, alors il est k-identifiable pour tout entier k vérifiant $1 \leq k \leq r$.*

Preuve. Soient a et b deux r-jumeaux ; supposons qu'ils soient $(r+1)$-séparés par un sommet c, avec par exemple $d(a,c) \leq r+1$ et $d(b,c) > r+1$. Par r-gémélité, on ne peut avoir $d(a,c) \leq r$ et donc nécessairement $d(a,c) = r+1$. Considérons une chaîne de longueur $r+1$ joignant a à c

$$a = v_0 v_1 v_2 \cdots v_r v_{r+1} = c.$$

On a alors $d(a,v_r) = r$ et donc $d(b,v_r) \leq r$; il s'ensuit que $d(b,c) \leq d(b,v_r) + d(v_r,c) = r+1$, d'où la contradiction.

□

2.5 Exemple

Reprenons l'exemple de l'introduction. Nous avons obtenu le code 1-identifiant de la figure 12.

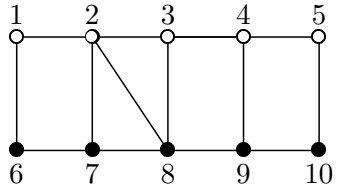


FIG. 12 – L'ensemble des sommets noirs constitue un code 1-identifiant de G

Pour savoir s'il est possible de trouver un code 2-identifiant dressons la liste, pour chaque sommet, des sommets qui le 2-couvrent.

	1	2	3	4	5	6	7	8	9	10
1	•	•	•			•	•	•		
2	•	•	•	•		•	•	•	•	
3	•	•	•	•	•		•	•	•	
4		•	•	•	•			•	•	•
5			•	•	•				•	•
6	•	•				•	•	•		
7	•	•	•			•	•	•	•	
8	•	•	•	•		•	•	•	•	•
9		•	•	•	•		•	•	•	•
10				•	•			•	•	•

Nous pouvons constater que ce graphe est bien sans 2-jumeaux et donc que le code $\mathcal{C} = \{1;2;3;4;5;6;7;8;9;10\}$ est 2-identifiant. Par contre, le choix précédent de $\mathcal{C} = \{6;7;8;9;10\}$ ne l'est plus. Nous laissons au lecteur intéressé le soin de déterminer un code 2-identifiant de ce graphe de taille minimum.

Enfin, il n'y a pas de code 3-identifiant sur ce graphe car les sommets 2 et 8 sont jumeaux : leur boule de rayon 3 est égale au graphe entier.

2.6 Une borne inférieure

Dans le cas général, nous n'avons guère de résultat meilleur que la proposition 1.

Proposition 7. *Soient $r \geq 1$ un entier, $G = (V, E)$ un graphe sans r-jumeaux et $\mathcal{C}$ un code r-identifiant de G. Alors*

$$|\mathcal{C}| \geq \lceil \log_2(|V| + 1) \rceil .$$

Preuve. Il suffit de remarquer que cette fois l'application injective

$$v \mapsto B(x, r) \cap \mathcal{C}$$

est à valeurs dans l'ensemble des parties non vides de $\mathcal{C}$, qui compte $2^{|\mathcal{C}|} - 1$ éléments. □

2.7 Puissances d'un graphe

Définition 8. *Soit $G = (V, E)$ un graphe et $k \geq 1$ un entier. La puissance $k^{ième}$ du graphe G est le graphe noté $G^{[k]}$, dont l'ensemble de sommets est V et tel que pour tous $x \neq y \in V$, l'arête xy est dans $G^{[k]}$ si et seulement si $d(x, y) \leq k$ dans G.*

On a alors le

Théorème 9. *Soit $G = (V, E)$ un graphe et $s, t \geq 1$ deux entiers. Soit $\mathcal{C} \subset V$. Alors*

$$\mathcal{C} \text{ est un code } st\text{-identifiant de } G \iff \mathcal{C} \text{ est un code } t\text{-identifiant de } G^{[s]}.$$

Preuve. Il suffit de voir que $c \in \mathcal{C}$ st-couvre un sommet x dans G si et seulement si c t-couvre x dans $G^{[s]}$. □

L'intérêt de ce résultat est qu'il permet en pratique d'obtenir un algorithme de recherche des codes r-identifiants si l'on dispose d'un algorithme de recherche des codes t-identifiants pour t diviseur de r ; en particulier pour $t = 1$.

D'un point de vue plus théorique, le problème Code 1-identifiant semble donc s'avérer plus difficile que le problème Code r-identifiant pour $r \geq 1$. Nous verrons à la prochaine section qu'ils sont en fait de difficulté équivalente puisque tous deux NP-complets.

2.8 Complexité du problème

Nous nous intéressons ici à la complexité algorithmique de deux problèmes : déterminer si un graphe est r-identifiable, et la détermination d'un code identifiant de cardinal minimum. Nous montrons que le premier est polynomial, alors que le second est NP-difficile. Dans cette section n désigne le nombre de sommets du graphe $G = (V, E)$.

r-identifiabilité
Instance : Un graphe $G = (V, E)$; un entier $r \geq 1$.
Question : G est-t-il sans r-jumeaux ?

Propriété 10. *Il existe un algorithme polynomial pour le problème r-Identifiabilité.*

Preuve. Il est possible de calculer la matrice des distances d'un graphe en temps $0(n^2)$, par un algorithme du type Dijkstra ; nous renvoyons à [5] pour l'exposé de cet algorithme. On peut alors déterminer en un temps $O(n)$ les boules $B(v, r)$ pour $v \in V$. Pour comparer deux parties de V, il faut au plus un temps $O(n)$; puisqu'il y a n boules à comparer entre elles, le tout est faisable (naïvement) en un temps majoré par $0(n^3)$ dans le pire des cas. □

Un problème plus difficile est celui-ci :

CODE 1-ID
INSTANCE : Un graphe $G = (V, E)$; un entier $k \geq 1$.
QUESTION : Existe-t-il un code identifiant de G de cardinal inférieur ou égal à k ?

Nous avons le résultat suivant :

Théorème 11. *Le problème* CODE 1-ID *est NP-complet.*

Ce résultat a été prouvé dans [4], par réduction du problème 3-SAT au problème Code 1-Id. Nous présentons ici un preuve originale, basée sur le problème Vertex Cover. Rappelons qu'un *ensemble couvrant* (les arêtes) dans un graphe $G = (V, E)$ est un sous ensemble $\mathcal{C} \subset V$ tel que pour toute arête $e = xy \in E$ on ait $x \in \mathcal{C}$ ou $y \in \mathcal{C}$. Autrement dit, un ensemble de sommets est couvrant si chaque arête du graphe lui est incidente. Le problème de décision associé VERTEX COVER admet la formulation qui suit :

VERTEX COVER
INSTANCE : Un graphe $G = (V, E)$; un entier $k \geq 1$.
QUESTION : Existe-t-il un ensemble couvrant de G de cardinal inférieur ou égal à k ?

Ce problème est NP-complet ; nous renvoyons à [8] pour la preuve de ce résultat, ainsi que pour le principe général des preuves de NP-complétude par réduction polynomiale. Nous pouvons maintenant démontrer le théorème 11.

Preuve. Etant donné un code $\mathcal{C} \subset V$ il est possible, en utilisant une procédure similaire à celle du théorème 10, de calculer en un temps $0(n^3)$ les ensembles $B(x, r) \cap \mathcal{C}$ et de les comparer. Le problème *Code 1-Id* est donc dans NP.

Pour prouver que ce problème est NP-complet nous considons un graphe $G = (V, E)$ à n sommets et m arêtes ; nous allons lui associer un graphe $G' = (V', E')$, constructible à partir de G en un temps polynomial, tel la propriété suivante soit vérifiée :

G admet un ensemble couvrant de cardinal inférieur ou égal à k si et seulement si G' admet un code 1-identifiant de cardinal inférieur ou égal à $k + 4n + 6m$.

Nous construisons G' en ajoutant à G des sommets et des arêtes de la manière suivante :

- pour tout $v \in V$ nous ajoutons une structure notée S_v composée de six sommets $v_1 \cdots v_6$ et six arêtes
$$vv_1,\ v_1v_2,\ v_2v_3,\ vv_4,\ v_4v_5,\ v_5v_6,$$
conformément à la figure 13 ;
- pour chaque arête $e = xy \in E$ nous ajoutons une structure S_e composée de onze sommets $e_1, e_2 \cdots e_{11}$ et treize arêtes
$$xe_1,\ ye_2,\ e_1e_2,\ e_1e_3,\ e_2e_3,\ e_3e_4,\ e_4e_5,\ e_3e_6,\ e_6e_7,\ e_7e_8,\ e_3e_9,\ e_9e_{10},\ e_{10}e_{11}$$
conformément à la figure 15. Cette structure restant inchangée si l'on inverse les rôles de x et y, peut importe le choix fait lors de la construction.

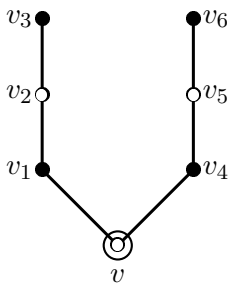


FIG. 13 – Structure S_v ajoutée sur chaque sommet v de G - les mots de code sont en noir

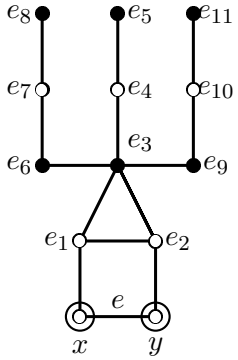


FIG. 14 – Structure S_e ajoutée sur chaque arête $e = xy$ de G - les mots de code sont en noir

Le graphe G' est donc constitué des n sommets et m arêtes de G, auxquels on a ajouté exactement les $6n + 11m$ sommets et $6n + 13m$ arêtes correspondant aux n structures S_v et m structures S_e ci-dessus ; il est donc clair que ce graphe peut être construit en un temps polynomial à partir de G.

Supposons pour commencer que G admette un ensemble couvrant $\mathcal{C} \subset V$ de cardinal inférieur où égal à k. Nous ajoutons à $\mathcal{C}$ d'autres sommets de manière à obtenir un code identifiant $\mathcal{C}'$ de G'. Plus précisément :

- pour chaque sommet $v \in V$ nous ajoutons sur S_v les quatre mots de code v_i pour $i \in \{1, 3, 4, 6\}$ (en noir sur la figure 13).
- pour chaque arête $e \in E$ nous ajoutons sur S_e les six mots de code e_i pour $i \in \{3, 5, 6, 8, 9, 11\}$ (en noir sur la figure 15).

L'ensemble $\mathcal{C}'$ obtenu a bien au plus pour cardinal $k + 4n + 6m$; montrons qu'il s'agit d'un code identifiant de G' :

- chaque sommet $v \in V$ est couvert par les mots de code $\{v_1, v_4\}$, ce qui l'identifie clairement ;
- dans chaque structure S_v, il est facile de voir qu'on a bien un code identifiant : par exemple v_1 est couvert uniquement par lui-même, v_3 est également couvert uniquement par lui-même et v_2 est couvert par v_1 et v_3. La branche $v_4v_5v_6$ est analogue.
- Dans chaque structure S_e, les sommets e_i sont clairement couverts et identifés par les mots de code choisis pour $i > 3$. Le sommet e_3 est le seul à être couvert par le triplet $\{e_3, e_6, e_9\}$, ce qui l'identifie. Reste à montrer que les sommets e_1 et e_2 sont couverts et identifiés. Ils sont tous deux couverts par e_3 ; pour les séparer, il est nécessaire qu'un des deux sommets x ou y au moins ait été choisi comme mot de code : ce qui est le cas puisque l'on a considéré au départ un ensemble couvrant de G.

Ces structures étant disjointes, chaque sommet de G' est bien couvert par un ensemble non-

vide de mots de codes qui lui est propre.

Réciproquement, supposons que $\mathcal{C}'$ soit un code identifiant de G' de cardinal inférieur ou égal à $k + 4n + 6m$ et montrons que $\mathcal{C} = \mathcal{C}' \cap V$ est un ensemble couvrant de G de cardinal inférieur ou égal à k.

- Considérons une structure S_v pour un $v \in V$. Pour séparer v_2 de v_3, le sommet v_1 a nécessairement été choisi comme mot de code. Pour couvrir v_3, il y a au moins un mot de code parmi v_2 et v_3. En faisant la même analyse dans la branche $v_4v_5v_6$, nous voyons qu'il y a au moins quatre mots de code sur S_v.
- Sur une structure S_e où $e \in E$, une analyse similaire sur les trois branches $v_3v_4v_5$, $v_6v_7v_8$ et $v_9v_{10}v_{11}$ montre qu'il y a au moins six mots de code.

Par conséquent, nous avons dans $\mathcal{C}'$ au moins $4n + 6m$ mots de code qui ne sont pas des sommets de G ; ainsi $\mathcal{C} = \mathcal{C}' \cap V$ est de cardinal inférieur ou égal à k.

Pour conclure, considérons une arête $e = xy$ de G. Dans S_e, les sommets e_1 et e_2 doivent être séparés par le code : ceci ne peut se faire que si au moins un des deux sommets x ou y est un mot de code, et donc appartient à $\mathcal{C}$. Ceci étant vrai pour tout $e \in E$, $\mathcal{C}$ est bien un ensembe couvrant.

□

Pour la détermination d'un code r-identifiant minimal dans un graphe quand $r > 1$, nous renvoyons à [3] où il est démontré que ce problème est NP-complet.

La recherche d'un code 1-identifiant apparaissant naturellement comme un restriction du problème DISCRIMINATION, nous sommes maintenant à même de prouver le théorème 2 :

Preuve. Le problème DISCRIMINATION est bien dans NP ; si n désigne le nombre de sommets de l'instance (individus+attributs), on peut aisément vérifier en un temps $O(n^2)$ qu'un code donné est discriminant. Soit maintenant $G = (V, E)$ un graphe ; on lui associe le graphe biparti $G' = (I \cup A, E')$ suivant :

- I et A sont deux copies de l'ensemble V des sommets de G ; on note leurs élements i_v (*individu* correspondant au sommet v) et a_v (*attribut* correspondant à la boule $B(v, 1)$) où $v \in V$.
- L'arête $i_v a_w$ est dans E' si et seulement si $v \in B(w, 1)$ dans G.

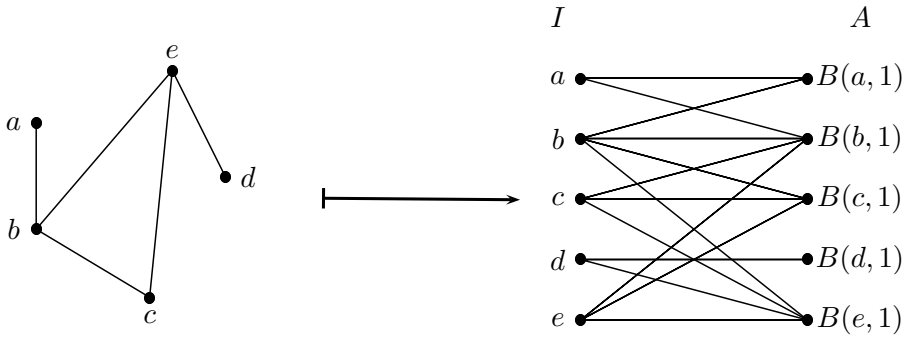


FIG. 15 – Les codes identifiants de G correspondent au codes discriminants de G'

Etant donné $i_v \in V$, il est clair que l'ensemble des attributs de i_v est $\{a_w,\ w \in B(v, 1)\}$; une partie $\mathcal{C}$ de V est donc un code 1-identifiant de G si et seulement si sa copie dans A

$$\{a_v,\ v \in \mathcal{C}\}$$

est un code discriminant de G'. Le problème DISCRIMINATION est donc équivalent au problème CODE 1-ID et donc NP-complet.

□

2.9 Résultats pour deux familles de graphes classiques

2.9.1 Chaînes

Pour $n \geq 1$, la chaîne P_n est le graphe constitué de n sommets $v_1, v_2, \cdots v_n$, dont les arêtes sont exactement les $v_i v_{i+1}$ pour $1 \leq i \leq n-1$; la chaîne P_1 étant le graphe comportant un unique sommet, et aucune arête (voir en exemple la chaîne P_8 sur la figure 16).

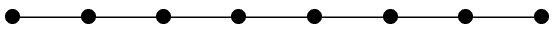


FIG. 16 – La chaîne P_8

Commençons par remarquer que :

Proposition 12. *Soient r et n deux entiers ≥ 1. La chaîne P_n est sans r-jumeaux si et seulement si $n \geq 2r+1$.*

Preuve. Montrons que si $n \geq 2r+1$ alors P_n est sans r-jumeaux ; la réciproque est une conséquence du théorème 18 dont la preuve figure en annexe. Appellons *frontière gauche* d'un sommet x une arête $v_i v_{i+1}$ telle que $v_i \notin B(x,r)$ et $v_{i+1} \in B(x,r)$ et *frontière droite* une arête $v_i v_{i+1}$ telle que $v_i \in B(x,r)$ et $v_{i+1} \notin B(x,r)$(voir figure 17).

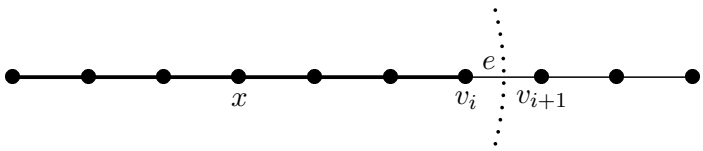


FIG. 17 – Cas $r = 3$. $e = v_i v_{i+1}$ est une frontière droite de x.

Si $e = v_i v_{i+1}$ est une frontière gauche de x alors nécessairement $x = v_{i+1+r}$; de même si $e = v_i v_{i+1}$ est une frontière droite de v alors $x = v_{i-r}$. Pour un sommet v quelconque de la chaîne P_n, la boule de centre v et de rayon r est une chaîne comportant au plus $2r+1$ sommets. Il est donc clair que si $n \geq 2r+1$ chaque sommet possède une frontière, sauf dans le cas $n = 2r+1$ où le sommet central v_{r+1} est le seul à ne pas en posséder.

Ceci considéré, si $n \geq 2r+1$, et si x et y sont deux sommets distintcs de la chaîne alors au moins un des deux possède une frontière $e = v_i v_{i+1}$, par exemple x. Supposons qu'il s'agisse d'une frontière gauche, l'autre cas étant similaire. Si $B(x,r) = B(y,r)$, alors $e = v_i v_{i+1}$ est aussi une frontière gauche de y ; on a alors $x = y = v_{i+1+r}$. Ceci prouve que ce graphe est sans r-jumeaux.

□

Nous supposons donc dorénavant que $n \geq 2r+1$ et allons considérer le cardinal d'un code identifiant de taille minimum sur P_n. Le résultat suivant a été prouvé dans [1], nous en donnons une preuve sensiblement différente.

Théorème 13. *Soient n et r deux entiers ≥ 1. Si $\mathcal{C}$ est un code r-identifiant sur P_n, alors*

$$|\mathcal{C}| \geq \left\lceil \frac{n+1}{2} \right\rceil .$$

Preuve. En reprenant la terminologie de la preuve de la propriété 12, nous allons ajouter des frontières imaginaires aux sommets ayant moins de deux frontières. Remarquons qu'un sommet x quelconque admet une frontière gauche si et seulement s'il ne couvre pas v_1, et une frontière droite si et seulement s'il ne couvre pas v_n.

Ajoutons deux arêtes imaginaires f_g et f_d au graphe sur les sommets v_1 et v_n, comme sur la figure 18. Redéfissons la notion de frontière par :

- une arête $e = v_i v_{i+1}$ de P_n est dite frontière d'un sommet x si exactement un des deux sommets v_i et v_{i+1} est dans la boule $B(x, r)$;
- l'arête imaginaire f_g est dite frontière d'un sommet x si v_1 est r-couvert par x ;
- l'arête imaginaire f_d est dite frontière d'un sommet x si v_n est r-couvert par x

Avec ces définitions, un sommet quelconque de la chaîne admet exactement deux frontières (éventuellement imaginaires) ; mais chacune de ces $n-1+2 = n+1$ arêtes doit être la frontière d'un sommet du code $\mathcal{C}$: en effet les sommets adjacents doivent être séparés et les sommets extrémaux v_1 et v_n doivent être couverts. Nous en déduisons qu'il existe au moins $\frac{n+1}{2}$ mots de code, soit

$$|\mathcal{C}| \geq \left\lceil \frac{n+1}{2} \right\rceil .$$

□

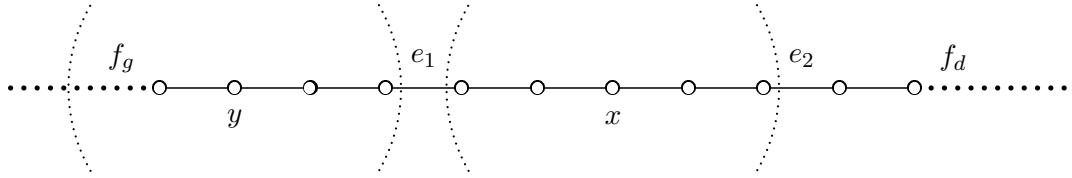


FIG. 18 – Cas $r = 2$. Les frontières du sommet x sont les arêtes e_1 et e_2 ; celles de y sont e_1 et l'arête "imaginaire" f_g.

Cette borne inférieure obtenue, nous cherchons maintenant à construire un code r-identifiant sur P_n de cardinal le plus petit possible. Commençons par le cas $r = 1$, en différenciant les cas suivant la parité de n.

Code 1-identifiant sur P_n, n impair ≥ 3.

On obtient un code 1-identifiant en considérant les $\frac{n+1}{2}$ mots de codes $v_1, v_3, \cdots v_n$ (voir fig. 19). En effet, les mots de codes sont couverts uniquement par eux mêmes, et les sommets v_i pour i pair sont couverts par les deux mots de codes v_{i-1} et v_{i+1}. Chaque sommet est donc bien identifié. Puisque n est impair on a $\frac{n+1}{2} = \left\lceil \frac{n+1}{2} \right\rceil$; donc d'après le théorème 13, ce code est optimal.

Code 1-identifiant sur P_n, n pair ≥ 3.

Pour n pair, la même construction semble fonctionner, à un détail près : les sommets v_{n-1} et v_n sont tous les deux couverts uniquement par le mot de code v_{n-1}. Pour les séparer, il nous suffit d'ajouter le mot de code v_{n-2} (voir fig. 20).

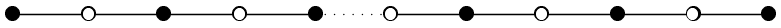


FIG. 19 – Code 1-identifiant sur P_n pour n impair. Les codes sont en noir.

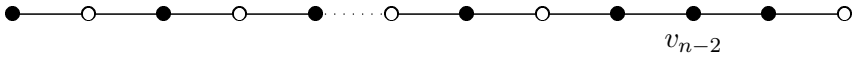


FIG. 20 – Code 1-identifiant sur P_n pour n pair. Les codes sont en noir.

Le code obtenu comporte alors $\frac{n}{2}+1$ sommets, mais comme n est pair on a $\frac{n}{2}+1 = \left\lceil \frac{n+1}{2} \right\rceil$. Ce code est encore optimal et on peut donc énoncer :

Théorème 14. *Soit $n \geq 3$. Les codes 1-identifiants de cardinal minimum sur P_n ont $\left\lceil \frac{n+1}{2} \right\rceil$ mots de code.*

Pour le cas $r > 1$ des constructions rencontrant la borne du théorème 13 existent pour une infinité de valeurs de n, mais le cas général est encore ouvert à l'heure actuelle.

2.9.2 Cycles

Pour $n \geq 1$, le cycle C_n est le graphe à n sommets $v_1, \cdots v_n$ dont les arêtes sont les $v_i v_{i+1}$ pour $1 \leq i \leq n-1$ ainsi que l'arête $v_n v_1$.

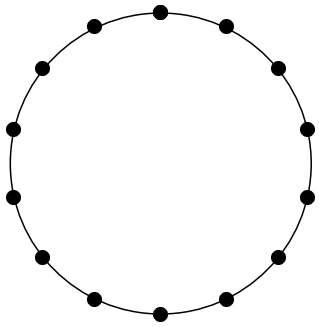


FIG. 21 – Le cycle à 14 sommets C_{14}

De part sa structure il est clair que le cycle C_n est sans r-jumeaux dès lors que $r < \mathrm{rad}(C_n)$. Le rayon de C_n étant $\left\lfloor \frac{n}{2} \right\rfloor$, on en déduit facilement que C_n est sans r-jumeaux si et seulement si $n \geq 2r+2$. On obtient facilement la borne suivante :

Théorème 15. *Soit $r \geq 1$ et $n \geq 2r+2$. Un code r-identifiant sur C_n a un cardinal supérieur ou égal à $\left\lceil \frac{n}{2} \right\rceil$.*

Preuve. On procède comme pour les chaînes. Dès lors que $n \geq 2r+2$ tout sommet admet v admet deux r-frontières, qui sont les arêtes séparant $B(v,r)$ de son complémentaire ; un mot de code permet donc de séparer exactement deux paires de sommets adjacents. Comme il y a n paires de sommets adjacents à sépaper on en déduit le résultat. □

Cette borne est optimale pour certaines valeurs de (n, r), en particulier dans le cas $n \geq 2r+4$ impair : nous renvoyons à [1] pour la preuve. Cependant, cette borne peut aussi être en deçà de la réalité, par exemple dans le cas extrémal $n = 2r + 2$:

Théorème 16. *Soit $r \geq 1$ et $n = 2r+2$. Les codes r-identifiants de C_n ont un cardinal supérieur ou égal à $n - 1$.*

Preuve. Notons $v_1, v_2, \cdots v_{2r+2}$ les sommets de C_{2r+2} ; le sommet diamétralement opposé à un sommet v_i est par définition le sommet v_{i+r+1} si $i \leq r + 1$ et v_{i-r-1} si $i > r + 1$. Il est aisé de constater pour un sommet v que tous les sommets du cycle sont dans la boule $B(v, r)$, à l'exception du sommet diamétralement opposé à v, que nous noterons v'. Clairement, $(v')' = v$.

Soient v et w deux sommets distincts du cycle. Un code r-identifiant doit séparer les sommets v' et w', par un mot de code ; ce mot de code est exactement dans une des deux boules $B(v', r)$ et $B(w', r)$, et donc aussi dans un des complémentaires de ces deux boules. Un des deux sommets v ou w doit donc être un mot de code. Ce raisonnement étant valable pour tous les couples de sommets distincts (v, w), on voit qu'il faut nécessairement au moins $n - 1$ mots de code.

□

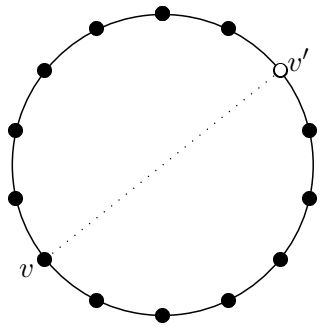


FIG. 22 – Cas $r = 6$: la boule $B(v, 6)$ (en noir) dans C_{14}

3 Quelques propriétés structurelles des graphes sans jumeaux

Rappelons que nous appelons *graphe sans r-jumeaux* un graphe simple non-orienté $G = (V, E)$ tel que pour tous $x, y \in V$,

$$x \neq y \implies B(x, r) \neq B(y, r).$$

Ces graphes sont précisément les graphes *r-identifiables*, c'est-à-dire admettant un code r-identifiant (voir la section 2 pour les différentes définitions). Il apparaît que cette propriété structurelle est intéressante en elle-même, et nous cherchons dans cette section à déterminer quels peuvent être les valeurs extr'males de quelques paramètres d'un graphe sans r-jumeaux.

3.1 Plus longues chaînes

Nous avons vu que la chaîne à $2r + 1$ sommets P_{2r+1} est la plus petite chaîne qui soit r-identifiable. D'autre part, le résultat suivant a été prouvé dans [2] :

Théorème 17. *Si G est un graphe r-identifiable, alors G contient une chaîne à $2r+1$ sommets en tant que graphe partiel.*

Ce résultat stipule donc l'existence dans un graphe r-identifiable $G = (V, E)$ de $2r + 1$ sommets $v_0, v_1 \cdots v_{2r}$ tels que $v_i v_{i+1} \in E$ pour tout $0 \leq i \leq 2r - 1$. Rien n'interdit cependant la présence d'une autre arête $v_i v_j \in E$ pour des valeurs non consécutives de i et j. Nous avons démontré le résultat suivant, répondant à une conjecture présente dans ce même article :

Théorème 18. *Si G est un graphe r-identifiable, alors G contient une chaîne à $2r+1$ sommets P_{2r+1} en tant que sous-graphe.*

La différence essentielle entre ces deux résultats est que la chaîne obtenue dans le théorème 18 est cette fois-ci *sans corde*. Notons l'existence du résultat suivant, dû à Erdös, Saks et Sós (voir [6] pour la preuve).

Théorème 19. *Soit G un graphe de rayon $\rho \geq 1$. Alors G contient une chaîne $P_{2\rho-1}$ en tant que sous-graphe .*

Soit $G = (V, E)$ un graphe de rayon ρ. Alors si v_0 est un centre de G et v_1 un voisin de v_0, on a $B(v_0, \rho + 1) = B(v_1, \rho + 1) = V$ et donc v_0 et v_1 sont $(\rho + 1)$-jumeaux. Ceci prouve le résultat suivant :

Un graphe r-identifiable a un rayon $rad(G) \geq r$; si $rad(G)=r$ alors G admet un unique centre.

Le théorème 19 implique donc l'existence dans tout graphe sans r-jumeaux d'une chaîne à $2r - 1$ sommets en tant que sous-graphe. Nous obtenons quant à nous le théorème 18 comme conséquence du résultat suivant :

Théorème 20. *Soit G un graphe de rayon ρ, et dont l'ensemble des centres constitue un stable. Alors G contient une chaîne $P_{2\rho+1}$ en tant que sous-graphe.*

3.2 Nombre de sommets

Disons quelques mots au sujet du nombre de sommets n d'un graphe sans r-jumeaux. Le théorème 18 nous fournit aisément une borne inférieure :

Proposition 21. *Soit $G = (V, E)$ un graphe sans r-jumeaux ; alors $|V| \geq 2r + 1$.*

Par contre, à r-fixé, il n'y a pas de borne maximale pour le nombre de sommets d'un graphe sans r-jumeaux : il considérer une chaîne P_k pour $k \geq 2r + 1$ quelconque.

3.3 Degré minimal

Rappelons que dans un graphe $G = (V, E)$, le *degré* $d(v)$ d'un sommet $v \in V$ est le nombre de sommets $w \in V$ tels que $vw \in E$. Le *degré minimal* du graphe G est alors simplement

$$\delta = \min_{v \in V} d(v).$$

Dès que $n \geq 2r+1$, la chaîne P_n constitue un graphe r-identfiable de degré minimal $\delta = 1$ et constitue donc une borne inférieure pour le degré minimal quand r et n sont fixés. A l'inverse, nous nous posés la question de savoir quel valeur maximale pouvait atteindre le degré minimal d'un graphe r-identifiable sur n sommets. Considérons le *graphe circulant* $\mathcal{C}^{[k]}_{2kr+2}$, puissance $k^{\text{ième}}$ du cycle à $2kr+2$ sommets $\mathcal{C}_{2kr+2}$ (voir la définition 8 pour la puissance $k^{\text{ième}}$). La figure 23 représente en exemple le graphe circulant $\mathcal{C}^3_{14}$ ($r = 2$ et $k = 3$).

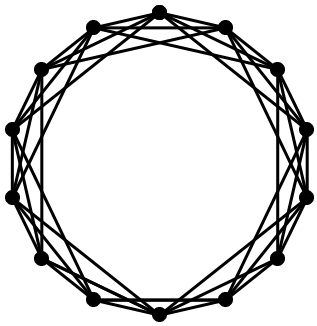


FIG. 23 – Le graphe circulant $\mathcal{C}^3_{14}$

Ce graphe est construit de telle manière qu'un sommet quelconque est précisément à distance $r+1$ du sommet diamétralement opposé, tous les autres sommets étant à distance inférieure ou égale à r. Il s'ensuit que $\mathcal{C}^{[k]}_{2kr+2}$ est r-identifiable, or son degré minimal est $\delta = 2k$. On a donc

$$\frac{n}{r} = \frac{2kr+2}{r} = \delta + \frac{2}{r}$$

soit

$$\delta = \left\lfloor \frac{n}{r} \right\rfloor$$

dès que $r > 2$. Nous conjecturons que ceci est la meilleure borne possible, plus précisément :

Conjecture 22. *Soit $r \geq 1$ et $G = (V, E)$ un graphe sans r-jumeaux, de degré minimal $\delta(G)$. Alors*

$$\delta(G) \leq \frac{n}{r}.$$

Malheureusement, à l'heure actuelle la meilleure borne qui ait été prouvée est (asymptotiquement)

$$\delta(G) \leq 1.5\frac{n}{r}.$$

Elle provient du résultat suivant, issu de [7] :

Théorème 23. *Soit G un graphe connexe sur n sommets et de degré minimum $\delta \geq 2$. Alors*

$$rad(G) \leq \frac{3}{2}\frac{n-3}{\delta+1} + 5$$

où $rad(G)$ désigne le rayon de G.

4 Deux autres types de codes métriques

Dans cette section nous présentons deux autres types de codes métriques : les codes *localisateurs-dominateurs* ainsi que les *codes de position*. Après les avoir définis nous donnons quelques exemples et étudions la complexité algorithmique de la recherche d'un code de cardinal minimum.

4.1 Codes localisateurs-dominateurs

Les codes localisateurs-dominateurs sont une variante des codes identifiants. Soient $G = (V, E)$ un graphe et $r \geq 1$ un entier. Une partie $\mathcal{C}$ de V est un code r-localisateur-dominateur si pour tous les sommets $x \neq y$ dans $V \setminus \mathcal{C}$ on a

$$B(x, r) \cap \mathcal{C} \neq B(y, r) \cap \mathcal{C}.$$

Ce qui différencie un code r-localisateur-dominateur d'un code r-identifiant est uniquement que l'on impose pas au mots de codes d'être identifiés ; on peut considérer qu'il le sont "automatiquement". Remarquons que pour tout graphe G, le code $\mathcal{C} = V$ convient trivialement ; tout graphe admet donc un code r-localisateur-dominateur.

Nous allons utiliser une preuve similaire à celle du théorème 11 pour montrer que le problème de la recherche d'un code 1-localisateur-dominateur est NP-complet. Formalisons le problème :

CODE 1-LOCALISATEUR-DOMINATEUR
INSTANCE : Un graphe $G = (V, E)$; un entier $k \geq 1$.
QUESTION : G admet-il un code 1-localisateur-dominateur de cardinal inférieur ou égal à k ?

Théorème 24. *Le probleme* CODE 1-LOCALISATEUR-DOMINATEUR *est NP-complet.*

Preuve. Il est une nouvelle fois simple de voir que ce problème fait partie de NP. Pour la NP-complétude, nous effectuons une réduction polynomiale du problème VERTEX COVER au problème CODE 1-LOCALISATEUR-DOMINATEUR. Soit $G = (V, E)$ un graphe à n sommets et m arêtes. Nous construisons un graphe G', constructible en temps polynomial à partir de G, tel que la propriété suivante soit vérifiée :

G admet un ensemble couvrant de cardinal inférieur ou égal à k si et seulement si G' admet un code 1-localisateur-dominateur de cardinal inférieur ou égal à $k + 2m$.

Pour tout $e = xy \in E$ nous ajoutons à G une structure S_e (voir fig. 24) constituée de 6 sommets notés e_1, e_2, e_3 et e'_1, e'_2, e'_3 et des arêtes

$$e_1e_2,\ e_1e_3,\ e_2e_3, e'_1e'_2,\ e'_1e'_3,\ e'_2e'_3$$

qui lui sont internes, ainsi que des arêtes

$$e_1x,\ e_1y,\ e_2x,\ e_2y,\ e'_1x,\ e'_1y,\ e'_2x,\ e'_2y$$

qui la joignent à l'arête e.

Supposons que G admette un ensemble couvrant $\mathcal{C}$ de cardinal inférieur ou égal à k. Nous ajoutons à $\mathcal{C}$ les mots de code e_1 et e'_1 pour tout $e \in E$ (en noir sur les deux figures), obtenant un ensemble $\mathcal{C}'$ de cardinal inférieur à $k + 2m$. Montrons que tous les sommets qui ne sont pas des mots de code sont identifiés :

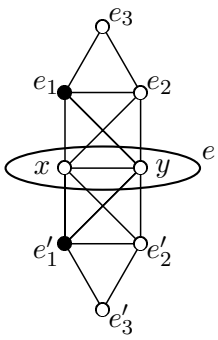


FIG. 24 – La structure S_e ajoutée sur l'arête $e = xy$

- sur chaque structure S_e, e_3 (respectivement e'_3) est couvert uniquement par e_1 (resp. e'_1), et e_2 (resp. e'_2) est couvert par e_1 (resp. e'_1) ainsi qu'un des deux sommets de l'arête e, puisque $\mathcal{C}$ est un ensemble couvrant. Les sommets de S_e qui ne font pas partie du code $\mathcal{C}'$ sont donc bien identifiés.
- Considérons une arête $e = xy$ avec $x, y \in V$. De ces deux sommets un au moins est un mot de code, par exemple y. Le sommet x est alors couvert (au moins) par e_1 et e'_1, ce qui le distingue de tous les autres sommets de G' qui ne sont pas des mots de code. A part dans le cas où G est réduit à un seul sommet, auquel cas il n'y a rien à démontrer, par connexité de G tout sommet $v \in V$ est incident à une arête ; s'il n'est pas un mot de code il sera donc identfifé pour les raisons précédentes.

$\mathcal{C}'$ est donc bien un code 1-localisateur-dominateur de G'. Réciproquement, supposons que $\mathcal{C}'$ soit un code 1-localisateur-dominateur de G' de cardinal inférieur ou égal à $k + 2m$. Sur chaque structure S_e, pour distinguer e_1 de e_2 il faut qu'un de ces deux sommets soit un mot de code ; de même pour e'_1 et e'_2. Par symétrie, nous pouvons supposer que e_1 et e'_1 sont dans $\mathcal{C}'$ pour tout $e \in E$.

Supposons maintenant qu'une arête $e = xy$ de G (x et y dans V) ne contienne pas de mot de code. Puisque dans la structure S_e les sommets e_2 et e_3 doivent être séparés, l'un d'eux est nécessairement un mot de code ; de même pour e'_2 et e'_3. La structure S_e contient donc au moins quatre mots de code ; il est possible d'obtenir un nouveau code 1-localisateur-dominateur $\mathcal{C}''$ de cardinal inférieur à celui de $\mathcal{C}'$ en supprimant les mots de code contenus dans S_e, et en les remplaçant par e_1, e_2 et x ou y. Ce nouveau code a l'avantage sur le précédent de couvrir l'arête couvre l'arête e de G. En faisant ceci pour tout $e \in E$, nous obtenons un code 1-localisateur-dominateur $\mathcal{C}''$ de G', de cardinal inférieur à celui de $\mathcal{C}'$, et qui couvre chaque arête de G. Notre analyse a montré que chaque structure S_e compte au moins deux mots de code ; puisque $|\mathcal{C}'| \leq k + 2m$ on en déduit que $\mathcal{C} = \mathcal{C}' \cap V$ est donc un ensemble couvrant de G dont le cardinal est inférieur ou égal à k .

□

4.2 Codes de position

La *dimension métrique* d'un graphe a été introduite en 1976 dans [10] ; on pourra trouver de nombreux résultats à ce sujet dans [12]. Nous en avons retrouvé quelques-un indépendamment, en lien avec les autres types de code déja présentés. Pour ces raisons nous conservons ici notre terminologie de "codes de position", pour une notion qu'on pourra trouver ailleurs sous le nom anglais de *resolving set*, et présentons ici une preuve de complexité.

Définition 25. *Un code de position dans un graphe $G = (V, E)$ est une partie $\mathcal{C}$ de V telle que pour tous $x, y \in V$*

$$x \neq y \implies \exists c \in \mathcal{C}, \quad d(x, c) \neq d(y, c).$$

De manière équivalente, on peut considérer qu'un sommet v du graphe est caractérisé par la donnée des distances $d(v, c)$ qui le séparent des mots de code $c \in \mathcal{C}$. Comme dans le cas des codes localisateurs-dominateurs, les mots de code $c \in \mathcal{C}$ sont "automatiquement" identifiés puisqu'ils sont les seuls sommets v à vérifier $d(v, c) = 0$. Tout graphe $G = (V, E)$ admet donc un code de position trivial, le code $\mathcal{C} = V$. Le problème qui nous intéresse ici est le suivant :

CODE DE POSITION
INSTANCE : Un graphe $G = (V, E)$; un entier $k \geq 1$.
QUESTION : G admet-il un code de position de cardinal inférieur ou égal à k ?

La *dimension métrique* de G est par définition le cardinal minimum d'un code de position de G. Les figures 25 et 26 montrent que la dimension métrique de la chaîne P_n est 1 pour $n \geq 1$ et que la dimension métrique du cycle C_n est 2 pour $n \geq 3$ (voir le 2.9 pour la définition de ces graphes).

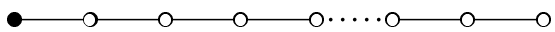


FIG. 25 – Un code de position minimum sur P_n

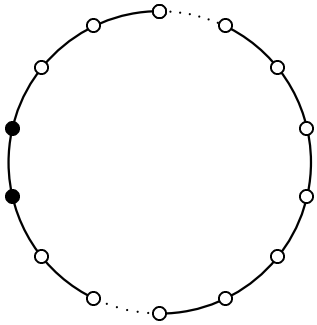


FIG. 26 – Un code de position minimum sur C_n

Malgré ces cas très simples, nous avons le résultat général suivant :

Théorème 26. *La détermination de la dimension métrique d'un graphe est un problème NP-complet.*

Preuve. Le calcul de la matrice des distances d'un graphe pouvant être effectué en temps polynomial, ce problème que nous appellerons CODE DE POSITION est dans NP. Soit maintenant $G = (V, E)$ un graphe et $k \geq 1$ un entier. Nous allons faire correspondre à G un graphe G', constructible en temps polynomial, vérifiant :

G admet un code localisateur-dominateur de cardinal $\leq k$ si et seulement G' admet un code de position de cardinal $\leq k + 1$

Le graphe $G' = (V', E')$ est défini par

$$V' = V \cup \{a, b, c\}$$

où a et b sont deux nouveaux sommets, et

$$E' = E \cup \{va,\ v \in V\} \cup \{ab\} \cup \{bc\}\,,$$

conformément à la figure 27.

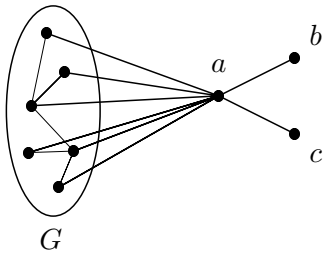


FIG. 27 – Le graphe G'

Supposons que G admette une code localisateur-dominateur $\mathcal{C}$ (à distance $r = 1$) de cardinal inférieur ou égal à k. Alors $\mathcal{C}' = \mathcal{C} \cup \{b\}$ est un code de position de G' de cardinal inférieur ou égal à $k + 1$, en effet :

- les mots de codes, en particulier b, sont identifiés par leur distance 0 à eux-même ;
- a est identifié comme étant le seul sommet à distance 1 de tous les mots de code ;
- c est identi fié comme étant le seul sommet à distance 2 de tous les mots de code ;
- les sommets de V qui ne sont pas des mots de codes sont identifiés par les codes qui leur sont adjacents, c'est-à-dire à distance 1.

Réciproquement, soit $\mathcal{C}'$ un code de position de G' de cardinal inférieur à $k + 1$. Alors nécessairement b ou c est un mot de code, car ces deux sommets sont à égale distance de tous les autres sommets ; par conséquent l'ensemble $\mathcal{C} = \mathcal{C}' \cap V$ est de cardinal inférieur à k : montrons qu'il s'agit bien d'un code localisateur-dominateur de G. Remarquons pour commencer que deux sommets distincts x et y, appartenant à V, sont à distance 1 s'ils sont adjacents et sinon sont à distance 2. Si x et y ne sont pas des mots de codes, alors il existe par définition $c \in \mathcal{C}'$ tel que $d(x, c) \neq d(y, c)$ dans G'. Par construction on voit que nécessairement $c \in V$ et donc que c est à distance 1 de x ou de y ; ces deux sommets sont donc séparés.

Pour que $\mathcal{C}$ soit un code-localisateur-dominateur de G, il faut encore que chaque sommet de V soit couvert par au moins un mot de code (dans V). Supposons que ce ne soit pas le cas : il existe alors (par séparation) exactement un sommet $x \in V$ qui n'est couvert par aucun mot de code ; mais alors pour séparer dans G' au sens des codes de position le sommet x des sommets a, b et c, il est nécessaire qu'il y parmi ces trois derniers un deuxième mot de code. En effet, dans le cas contraire, on peut supposer que l'unique mot de code parmi eux est b, et alors x et c ne sont plus séparés au sens des codes de position : ils sont à distance 2 de tous les mots de code. Nous voyons alors que $\mathcal{C}$ est de cardinal inférieur à $k - 1$, et donc que $\mathcal{C} \cup \{x\}$ est un code localisateur-dominateur de G de cardinal inférieur ou égal à k.

□

Annexe : Preuve du théorème 20

Nous rappelons quelques définitions avant de commencer la preuve. Soit $G = (V, E)$ un graphe connexe et $W \subset V$. Le sous-graphe de G induit par W est le graphe $G[W]$ dont l'ensemble des sommets est W et dont les arêtes sont les $e \in E$ ayant leur deux extrémités dans W. Si $v \in V$ on notera $G[V - v]$ pour $G[V \setminus \{v\}]$. Si G est connexe, un *sommet d'articulation* de G est un sommet $v \in V$ tel que $G[V - v]$ n'est pas connexe. On notera $\mathcal{C}(G)$ l'ensemble des sommets d'articulation de G, et $\mathcal{H}(G)$ son complémentaire. Ainsi, $v \in \mathcal{H}(G)$ si et seulement si $G[V - v]$ est connexe.

Nous allons démontrer le résultat suivant :

Théorème 1. *Soit G un graphe connexe de rayon r, et dont l'ensemble des centres constitue un stable. Alors G contient une chaîne P_{2r+1} en tant que sous-graphe.*

Il a été montré à la section 3.1 que ce théorème implique :

Corollaire 2. *Si G est un graphe sans r-jumeaux, alors G contient une chaîne à $2r+1$ sommets P_{2r+1} en tant que sous-graphe.*

Dans les deux résultats précédents, les graphes sont supposés avoir au moins deux sommets. Nous allons faire une hypothèse plus précise pour travailler et notons (Q_r) la propriété ci-dessous :

$$G = (V, E) \text{ est connexe, } |V| \geq 2 \text{ et} \qquad (Q_r)$$

$$\forall e = xy \in E \qquad B(x, r) \neq V \text{ ou } B(y, r) \neq V.$$

Autrement dit, un graphe connexe possédant au moins deux sommets vérifie (Q_r) si et seulement si sont rayon est strictement supérieur à r, ou si son rayon est r et que deux centres ne sont jamais adjacents. Nous démontrons ci-dessous que si $r \geq 1$ et que le graphe $G = (V, E)$ vérifie (Q_r), alors il existe $W \subset V$ tel que $G[W]$ soit une chaîne à $2r + 1$ sommets. Commençons par un lemme :

Lemme 3. *Soit $G = (V, E)$ un graphe connexe et $G' = G[W]$ un sous-graphe induit connexe de G. Alors pour tout $x, y \in W$*

$$d_{G'}(x, y) \geq d_G(x, y).$$

Preuve. Soit P une plus courte chaîne joignant x à y dans G', comportant k arêtes ; alors P constitue une chaîne joignant x à y dans G d'où

$$d_G(x, y) \leq k = d_{G'}(x, y).$$

□

Passons à la preuve du théorème 1.

Preuve. Si un graphe $G = (V, E)$ admet sous-graphe G' vérifiant (Q_r), l'existence pour G' d'une partie W telle que $G'[W]$ soit une chaîne à $2r + 1$ sommets P_{2r+1} impliquera par transitivité de la notion de sous-graphe le même résultat pour G. Nous pouvons donc nous contenter de prouver le résultat pour un graphe $G = (V, E)$ *minimal* parmi les graphes vérifiant (Q_r), en supposant précisément :

(*i*) G vérifie (Q_r) ;
(*ii*) si $W \subset V$ et $G[W]$ vérifie (Q_r) alors $W = V$.

Le théorème est vrai pour $r = 1$. En effet si G vérifie (Q_1), G possède au moins deux sommets donc aussi une arête $e = xy$ par connexité ; par (Q_1) il existe alors $z \in V$ tel que $d(x, z) > 1$ (ou $d(y, z) > 1$). Un chaîne de longueur minimale de x à z (ou de y à z) contient donc au moins trois sommets et donc un P_3, sous-graphe de G.

Supposons désormais $r \geq 2$. Nous allons montrer que G admet la structure suivante (fig. 28) :

il existe une partition de V *en* $k \geq 2$ *parties non-vides* P_1, P_2, $\cdots P_k$ *telle que*
- *pour tout* $1 \leq i \leq k$, $G[P_i]$ *est une chaîne* $a_i \cdots b_i$ *comportant entre* 2 *et* r *sommets ;*
- *les arêtes de* G*, hormis les arêtes des chaînes* $G[P_i]$*, sont toutes de la forme* $b_i b_j$.

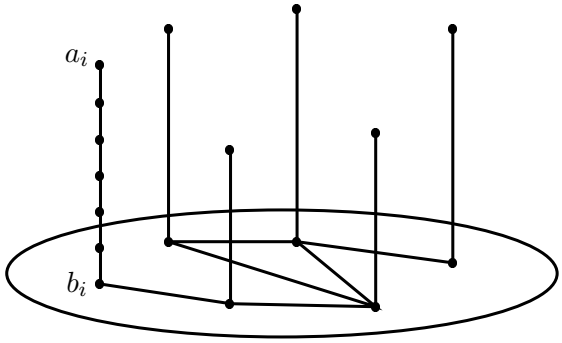


FIG. 28 – Structure d'un graphe (Q_r)-minimal

Etape 1 : sommets critiques de G

Par minimalité de G quel que soit $h \in \mathcal{H}(G)$, le sous-graphe $G[V - h]$ ne vérifie pas (Q_r). Voyons quelle hypothèse peut être mise en défaut :
- par définition de $\mathcal{H}(G)$, $G[V - h]$ est connexe ;
- puisque G vérifie (Q_r) il existe dans G au moins deux sommets, et donc une arête par connexité ; l'une des extrémités de cette arête admet par (Q_r) un sommet à distance au moins $r + 1 \geq 3$; donc G compte au moins quatre sommets. $G[V - h]$ contient donc au moins trois sommets, donc plus de deux.
- Pour contredire (Q_r), c'est donc qu'il existe donc $e = xy \in E$ telle que $B'(x, r) = B'(y, r) = V \setminus \{h\}$, les notations B' étant relatives à G'.

Par le lemme 3 on en déduit que dans le graphe de départ on a

$$B(x, r) \setminus \{h\} = B(y, r) \setminus \{h\} = V \setminus \{h\}.$$

Ceci signifie que dans G seul le sommet h peut être à distance strictement supérieure à r de x ou y ; et c'est nécessairement le cas pour au moins un des deux puisque G vérifie (Q_r). On a donc pour l'un des deux - par exemple pour x -

$$d(x, h) = r + 1 \text{ et } B(x, r) \setminus \{h\} = V \setminus \{h\}.$$

Nous appellerons un tel x *sommet critique de* h et noterons $\mathrm{crit}(h)$ l'ensemble des sommets critiques de h. Plus précisément,

$$\mathrm{crit}(h) = \{x \in V,\ d(x, h) = r + 1 \text{ et } \forall z \neq h \ \ d(x, z) \leq r\}.$$

Nous venons de montrer que

$$\textit{crit}(h) \textit{ est non vide pour tout } h \in \mathcal{H}(G). \tag{1}$$

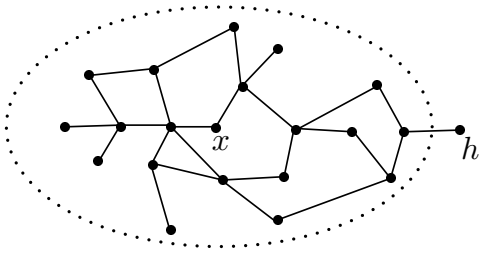


FIG. 29 – Cas $r = 4$; x est un sommet critique de h, qui est le seul sommet à ne pas être dans $B(x, 3)$

Nous allons maintenant montrer que l'ensemble critique d'un $h \in \mathcal{H}(G)$ ne peut être composé que de sommets d'articulation de G. Supposons pour cela l'existence de h_1 et $h_2 \in \mathcal{H}(G)$ tels que $h_1 \in \text{crit}(h_2)$. On a alors

$$d(x, h_1) \geq r + 1 \Longleftrightarrow x = h_2. \tag{2}$$

Or par (1), $\text{crit}(h_1)$ est non-vide. Mais un sommet $x \in \text{crit}(h_1)$ vérifie $d(x, h_1) = r+1$ et donc $x = h_2$ par (2) ; ainsi $\text{crit}(h_1) = \{h_2\}$, et par le même argument $\text{crit}(h_2) = \{h_1\}$. Les sommets h_1 et h_2 sont donc à distance $r + 1$ l'un de l'autre, et pour tout autre sommet

$$\text{si } v \in V \setminus \{h_1, h_2\} \text{ alors } d(h_i, v) \leq r \text{ pour } i = 1, 2. \tag{3}$$

Posons alors $G' = G[V \setminus \{h_1, h_2\}]$. Par (Q_r)-minimalité de G, G' ne vérifie pas (Q_r) : nous allons montrer que c'est parce que G' n'est pas connexe. En effet, nous avons déja établi que G possède au moins quatre sommets, donc G' en a bien au moins deux. Supposons que G' soit connexe ; alors pour contredire (Q_r) il doit exister une arête $e = xy \in E$ - où x, y sont distincts de h_1 et h_2 - telle que $B'(x, r) = B'(y, r) = V \setminus \{h_1, h_2\}$, les notations B' étant relatives à G'. Par le lemme 3 on en déduit que dans G on a $d(x, v) \leq r$ et $d(y, v) \leq r$ pour tous $v \neq h_1, h_2$. Avec (3) ceci implique que $B(x, r) = B(y, r) = V$, contredisant le fait que G vérifie (Q_r).

Ainsi, $G' = G[V \setminus \{h_1, h_2\}]$ ne peut être connexe et possède donc au moins deux composantes connexes (non-vides) C et C'. Puisque $G[V \setminus \{h_2\}]$ est connexe, h_1 possède au moins un voisin dans C et un voisin dans C' ; de même pour h_2. Il existe ainsi une plus courte chaîne P de h_1 à h_2 parmi les chaînes ne contenant, hormis leurs extrémités h_1 et h_2, que des sommets de C ; et de même une plus courte chaîne de h_1 à h_2 passant par C' (fig. 30).

Puisque $d(h_1, h_2) = r+1$, on obtient un cycle $h_1 P h_2 P' h_1$ à au moins $2r+2$ sommets qui est un sous-graphe de G ; il ne peut y avoir d'autres arêtes dans G joignant deux sommets du cycle, puisqu'il est composé de deux chaînes minimales, dans deux composantes disjointes, séparées par les sommets h_1 et h_2. Ce cycle contient strictement une chaîne à $2r + 1$ sommets, sous-graphe de G, qui vérifie (Q_r) et contredit la minimalité de G.

Nous avons donc démontré :

$$\textit{pour tout } h \in \mathcal{H}(G) \textit{ on a } \emptyset \neq \textit{crit}(h) \subset \mathcal{C}(G).$$

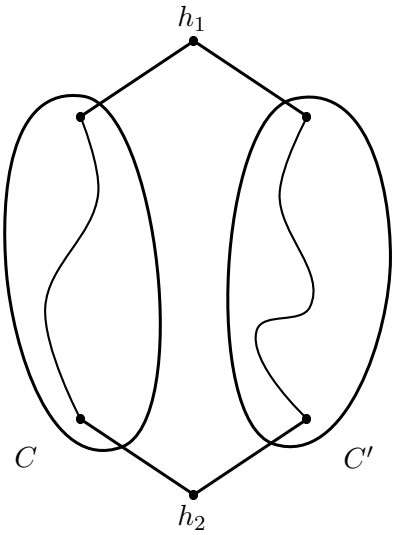


FIG. 30 – Le cas $\text{crit}(h_1) \subset \mathcal{H}(G)$

En particulier, puisque pour tout graphe $\mathcal{H}(G) \neq \emptyset$, on a $\mathcal{C}(G) \neq \emptyset$.

Etape 2 : décomposition autour d'un sommet d'articulation

Considérons maintenant $c \in \mathcal{C}(G)$ et soient C_1, C_2, ... C_k les k composantes connexes de $G[V - c]$ ($k \geq 2$).

Cas 1 : s'il existait au moins deux composantes C_i, par exemple C_1 et C_2, contenant chacune un x_i tel que $d(x_i, c) \geq r$, alors alors en considérant des plus courtes chaînes P_1 et P_2 joignant respectivement x_1 et x_2 à c, on obtiendrait une chaîne $P_1 c P_2$, sans corde, comportant au moins $2r + 1$ sommets (fig. 31).

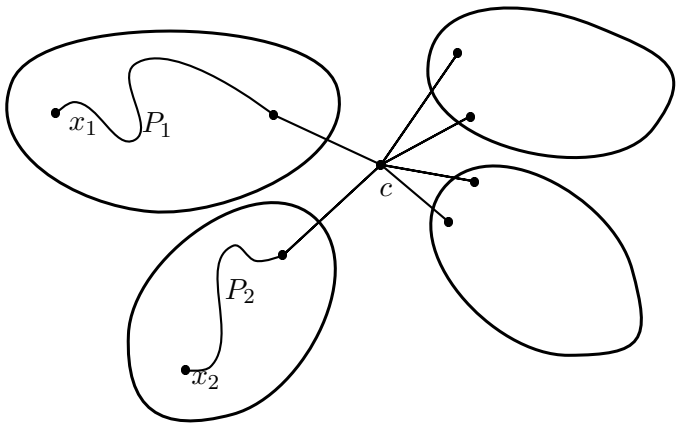


FIG. 31 – Etape 2, cas 1

G contiendrait donc P_{2r+1} et par minimalité on aurait $G = P_{2r+1}$; la conclusion du théorème en découlerait donc.

Cas 2 : supposons maintenant qu'il n'existe au plus qu'une composante, soit C_1, contenant des sommets dont la distance à c est supérieure ou égale à r. Soit $P = ca_1a_2 \cdots a_l$ une plus

longue chaîne joignant c à un sommet d'une composante C_i pour $i \geq 2$; on a alors $1 \leq l \leq r-1$. Montrons que $G' := G[C_1 \cup P]$ vérifie Q_r (fig. 32).

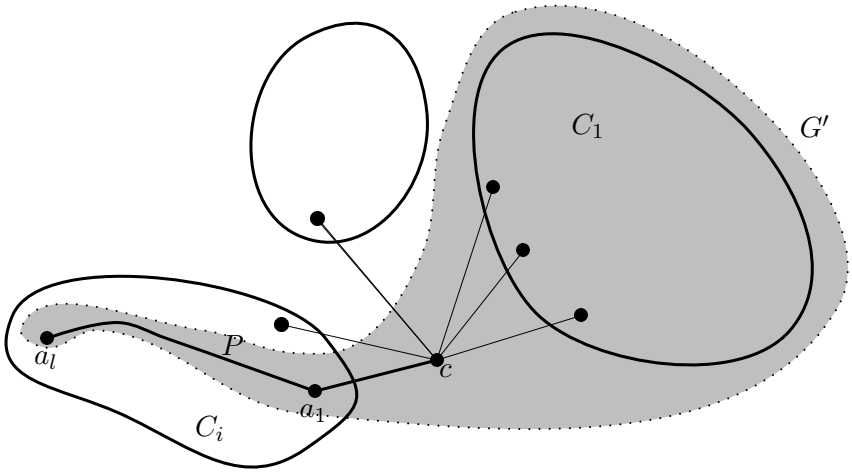


FIG. 32 – Etape 2, cas 2

Il est clair que G' contient une arête et est connexe ; soit $e = x_1x_2$ une arête de G'. Nous allons distinguer les cas pour montrer qu'il existe dans G' pour $i = 1$ ou $i = 2$ un sommet z vérifiant $d(x_i, z) \geq r + 1$, ce qui prouvera que G' vérifie (Q_r).

- Si x_1 et x_2 sont dans $C_1 \cup \{c\}$: puisque G vérifie (Q_r) il existe dans V un sommet z tel que $d_G(x_i, z) \geq r + 1$ pour $i = 1$ ou 2.
  - Si $z \in C_1$ ou $z \in P$ alors z est un sommet de G' et $d_{G'}(x_i, z) \geq d_G(x_i, z) \geq r + 1$.
  - Si $z \in C_j$ pour $j \neq 1$, alors $1 \leq d(c, z) < l$. Il existe donc sur P un sommet a_j tel que $d(c, a_j) = d(c, z)$. Les plus courtes chaînes de x_1, x_2 à z ou a_j passant toutes par c, on en déduit que $d_{G'}(x_i, a_j) = d_G(x_i, z) \geq r + 1$ (figure 33).

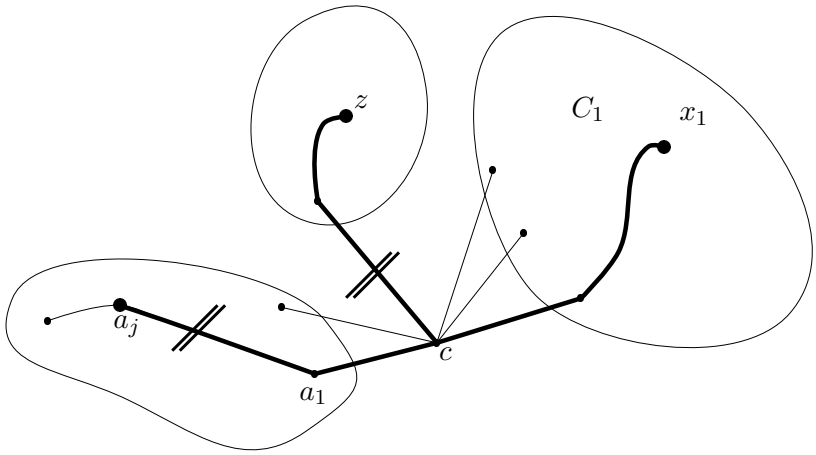


FIG. 33 – Etape 2, cas 2, premier tiret. Les distances cz et ca_j étant égales, on a $d(x_1, z) = d(x, a_j)$

- Considérons l'arête ca_1. Il existe un $z \in V$ tel que $d(c, z) \geq r + 1$ ou $d(a_1, z) \geq r + 1$; dans les deux cas $d(c, z) \geq r$ donc par hypothèse on ne peut avoir que $z \in C_1$ et donc z est un sommet de G' ; comme précédemment, la propriété (Q_r) est donc vérifiée pour l'arête ca_1.

– si maintenant on considère une arête $a_i a_{i+1}$, le sommet z obtenu dans le cas précédent conviendra aussi à cette arête.

Finalement, puisque G' vérifie (Q_r) on a par minimalité $G' = G$. Nous avons démontré le résultat suivant :

si $c \in \mathcal{C}(G)$ alors $G[V - c]$ admet exactement deux composantes connexes dont l'une, composée de l sommets $a_1, a_2, \cdots a_l$ avec $1 \leq l \leq r-1$, est telle que $a_1 \cdots a_l$ soit une chaîne. (fig. 34). (4)

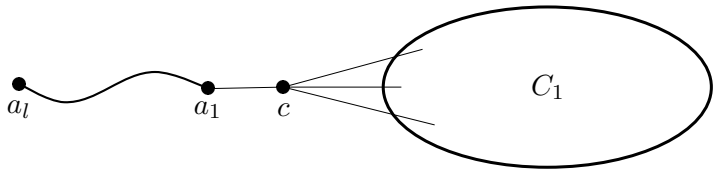


FIG. 34 – Décomposition autour d'un sommet d'articulation

Etape 3 : décomposition globale.
A partir de cette décomposition locale, valable en chaque sommet d'articulation de G, nous allons obtenir la décomposition globale qui suit. Il sera montré dans une prochaine étape qu'en fait $H_1 = \emptyset$, de manière à obtenir la structure évoquée en début de preuve.

Il existe une partition de V en $k+1$ ensembles $V = H_1 \cup P_1 \cup P_2 \cdots \cup P_k$ (où H_1 est le seul ensemble éventuellement vide) avec $k \geq 2$ où :
- *$H_1 \subset \mathcal{H}(G)$;*
- *Pour tout $1 \leq i \leq k$, $G[P_i]$ est une chaîne $a_i \cdots b_i$ à m_i sommets avec $2 \leq m_i \leq r$;*
- *Les arêtes de G, excepté les arêtes des chaînes $G[P_i]$, ont leur extrémités dans $H_1 \cup \{b_i, 1 \leq i \leq k\}$.*

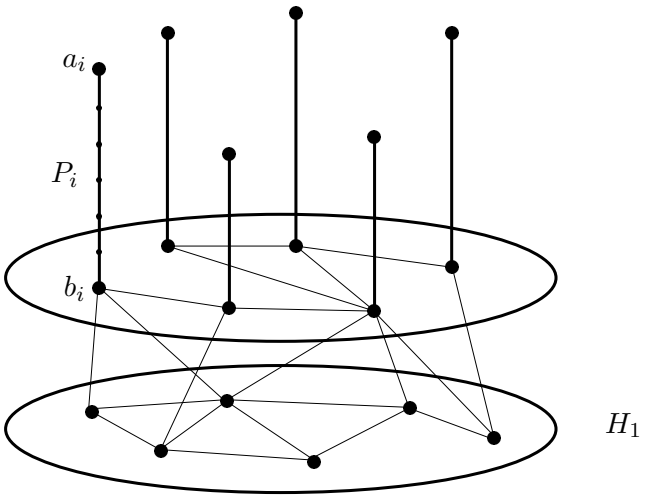


FIG. 35 – Etape 5

Puisqu'il existe au moins un sommet d'articulation dans G, on peut lui appliquer le résultat de décomposition locale (4). Cela montre l'existence d'une chaîne à l sommets $a_1 a_2 \cdots a_l$ dans G, avec $1 \leq l \leq r-1$, telle que a_1 est de degré 1 et les autres de degré 2 (nous inversons ici la numérotation). Remarquons aussi que $a_1 \in \mathcal{H}(G)$ et que $a_i \in \mathcal{C}(G)$ pour $i > 1$.

Pour les besoins de cette étape, définissons donc un *filament* comme étant une chaîne $a_1a_2\cdots a_l$ de G, avec $1 \leq l \leq r-1$, où les sommets sont de degré 2 à l'exception du premier, de degré 1. Nous dirons que le filament est *maximal* s'il n'existe pas de sommet a_{l+1} de degré 2 pouvant le prolonger. Clairement, tout filament est contenu dans un unique filament maximal.

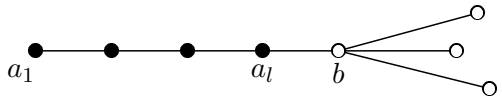


FIG. 36 – Un filament maximal (sommets noirs)

Etant donné un filament maximal $a_1a_2\cdots a_l$, le dernier sommet a_l admet un deuxième voisin b qui ne fait pas partie du filament : si ce voisin est de degré 1, alors G est une chaîne. Nous supposons dorénavant que G n'est pas une chaîne : alors le voisin b de a_l est un sommet de degré supérieur ou égal à 3. De plus, la suppression de b déconnecte le filament $a_1a_2\cdots a_l$ des autres voisins de b ; on voit donc que $b \in \mathcal{C}(G)$. Remarquons aussi que deux filaments maximaux P_1 et P_2 ne peuvent être rattachés à un même sommet b de degré supérieur à 3 ; en effet puisque $b \in \mathcal{C}(G)$, le lemme 4 implique que $G[V-b]$ n'a que deux composantes connexes. Nous avons donc obtenu les résultat suivant :

Il existe dans G des chaînes disjointes $P_i = a_i \cdots b_i$ comportant entre 2 et r sommets, où a_i est degré 1, b_i de degré 3 et les autres sommets de degré 2 ; et il y a unicité de cette décomposition.

Or tout sommet d'articulation est sur une des chaînes ci-dessus par (4) : soit il est sur un filament maximal, soit il s'agit d'un des sommets b_i. Pour obtenir la décomposition annoncée en début d'étape, reste à considérér l'ensemble H_1 des éléments de $H \in \mathcal{H}(G)$ qui n'ont pas été obtenus comme un sommet de degré 1 dans les filaments. De part la répartition des degrés sur les chaînes P_i, les voisins de H_1 ne peuvent être que des sommets b_i.

Etape 4 : $H_1 = \emptyset$.

Nous allons maintenant montrer que dans la décomposition de l'étape 4, on a toujours $H_1 = \emptyset$. Les extrémités a_i des chaînes P_i ne sont pas des sommets d'articulation de G, donc comme tous les éléments de $\mathcal{H}(G)$ admettent des sommets critiques ; nous allons montrer que ces sommets critiques ne peuvent être que des sommets b_j, autres extrémités des chaînes.

Soient donc $1 \leq i \leq k$ et $x \in \mathrm{crit}(a_i)$. Nous avons $x \in \mathcal{C}(G)$ donc x est sur l'une des chaînes P_j pour $j \neq i$ - en effet la longueur de P_i est strictement inférieure à r (moins de r sommets). Supposons que $x \neq b_j$ et considérons une plus courte chaîne minimale de a_i à x. Cette chaîne contient obligatoirement b_i et a_j ; soit a le sommet précédant b_j sur cette chaîne : nous allons montrer que l'arête ab_j contredit la propriété (Q_r). En effet, soit z un sommet quelconque de G :

- si $z \in P_j$ alors $d(a_j, z) < r$ et donc $d(a, z) \leq r$;
- si $z \notin P_j$ alors un chemin minimal de x à z passe toujours par a_j. Alors si $z \neq a_i$ on a $d(x, z) \leq r$ car $x \in \mathrm{crit}(a_i)$, d'où

$$d(a_j, z) < d(x, z) \leq r$$

donc aussi

$$d(a, z) \leq d(a_j, z) + 1 \leq r \ ;$$

– enfin, si $z = a_j$ alors puisque a_j et a sont sur une plus courte chaîne joignant a_i à x leur distance à a_i est inférieure à r. On a donc

$$B(a, r) = B(a_j, r) = V$$

d'où une contradiction.

Les $x \in \text{crit}(a_i)$ ne peuvent donc être que des sommets b_j pour $j \neq i$.

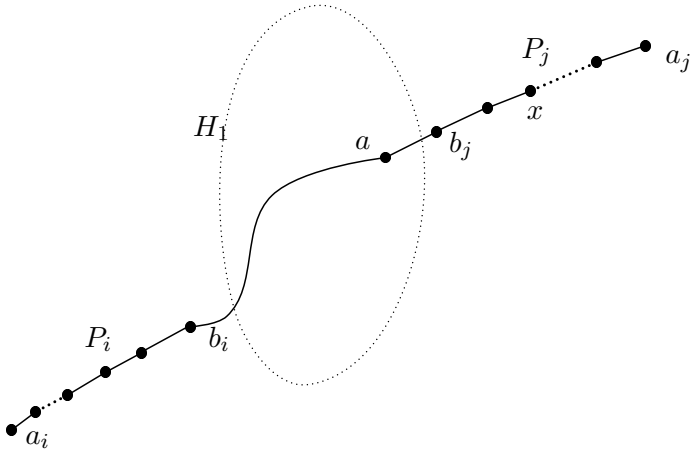


FIG. 37 – Etape 4

Remarquons que si $h_1 \neq h_2$ alors $\text{crit}(h_1)$ et $\text{crit}(h_2)$ sont disjoints. En effet si $x \in \text{crit}(h_1)$, x possède un unique point à distance $r+1$ qui est h_1 ; d'où l'identification de h_1. D'après ce qui précède, les $\text{crit}(a_i)$ étant non-vides et puisqu'il y a autant de a_i que de b_i, il y a bijection entre les ensembles $\text{crit}(a_i)$ et les singletons $\{b_j\}$.

Nous concluons alors en montrant que s'il existait $h \in H_1$, $\text{crit}(h)$ devrait aussi contenir des éléments b_j ; or les b_j sont tous "pris" par les a_i. Supposons donc l'existence de $h \in H_1$ et d'un $x \in \text{crit}(h)$, $x \neq b_j$ pour tout j donc au strictement au milieu d'une chaîne P_j (strictement entre b_j et a_j). Il existe i tel que $\text{crit}(a_i) = \{b_j\}$. Puisque $h \neq a_i$, on a $d(a_i, x) \leq r$ et donc il s'ensuit que $d(a_i, b_j) \leq r$: contradiction.

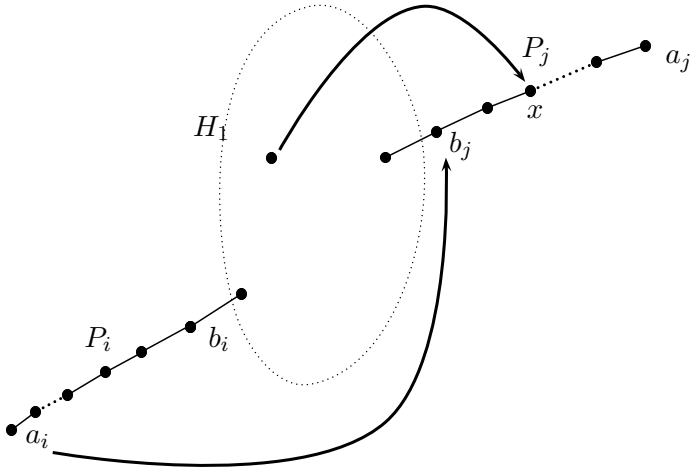


FIG. 38 – Etape 4. Les flèches indiquent les éléments critiques.

Nous avons montré que G est composé de $k \geq 2$ chaînes $P_i = a_1 \cdots b_i$, comportant chacune entre 2 et r sommets, ces chaînes n'étant reliées entre elles que par des arêtes entre les extremités b_i.

Etape 5 : récurrence

Nous pouvons alors procéder par récurrence à la démonstration du théorème ; le cas $r = 1$ a déjà été traité. Supposons alors $r \geq 2$ et considérons la décomposition obtenue par les propositions précédentes pour un graphe G qui soit (Q_r)-minimal.

Nous définissons le graphe $G' = G[V \setminus \{a_i, 1 \leq i \leq k\}]$: il est obtenu à partir de G en effaçant le sommet terminal de chaque chaîne P_i. Nous montrons que G' vérifie la propriété (Q_{r-1}) :

- G' est connexe ;
- G' admet au moins deux sommets car il y a au moins deux sommets b_i dans la décomposition (même dans le cas dégénéré où G est une chaîne) ;
- enfin, soit $x \neq a_i$ un sommet de G'. S'il existe dans G un sommet z tel que $d(x, z) \leq r+1$, alors c'est toujours le cas dans G', sauf si $z = a_i$; mais dans ce cas le prédécesseur de a_i sur la chaîne P_i vérifie $d(x, z) \leq r$.

Puisque G' vérifie Q_{r-1}, par récurrence il existe $W \subset V \setminus \{a_i, 1 \leq i \leq k\}$ tel que $G'[W]$ soit est une châine à $2r-1$ sommets $Q = w_0 w_1 \cdots w_{2r-2}$. Les extremités w_0 et w_1 sont nécessairement sur des chaînes P_i distinctes car la longueur des P_i est inférieure à r. Il s'ensuit qu'en revenant au graphe initial G (en ajoutant les a_i) on peut prolonger la chaîne Q sur la chaînes P_i contenant w_0 et w_{2r-1} ; les sommets ajoutés étant strictement à l'intérieur des P_i il ne peut y avoir de corde. Nous obtenons ainsi une chaîne sans cordes à $2r+1$ sommets, sous-graphe de G. Cette chaîne vérifiant elle-même (Q_r), par minimalité G *est* une chaîne à $2r+1$ sommets.

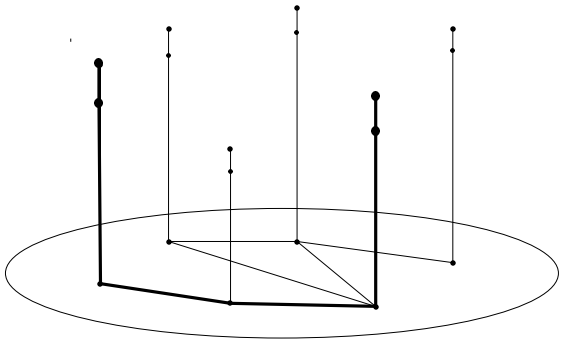


FIG. 39 – de P_{2r-1} à P_{2r+1}

□

Références

[1] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, *Identifying and locating-dominating codes on chains and cycles* European Journal of Cominatorics 25, pp. 969-987, 2004.

[2] I. Charon, H. Honkala, O. Hudry, A. Lobstein, *Structural Properties of Twin-Free Graphs*, soumis à Electronic Journal of Combinatorics, (2006).

[3] I. Charon, O. Hudry, A. Lobstein, *Identifying and Locating-Dominating Codes : NP-Completeness Results for Directed Graphs*, IEEE Transactions on Information Theory, Vol. IT-48, pp. 2192-2200, 2002.

[4] I. Charon, O. Hudry, A. Lobstein, *Minimizing the size of an Identifying or Locating-Dominating Code in a Graph is NP-hard*, Theoretical Computer Science, 290(3), pp.2109-2120, 2003.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction à l'algorithmique*, seconde edition Dunod, Paris, 2004.

[6] P. Erdös, M. Saks, V. T. Sós, *Maximum Induced Trees in Graphs* Journal of Combinatorial Theory, Series B41, pp.61-79, 1986.

[7] P. Erdös, J. Pach, R. Pollack, Z. Tuza, *Radius, Diameter ans Minimum Degree*, Journal of Combinatorial Theory, Series B47, pp. 73-79, 1989.

[8] M.R. Garey, D.S. Johnson, *Computers and Intractability : a guide to the theory of NP-Completeness*, New York : Freeman, 1979.

[9] S. Gravier, J. Moncel, A. Semri, *Identifying Codes on Cycles*, European Journal of Combinatorics, 27(5), pp. 767-776, 2006.

[10] F. Harary and R. A. Melter, *On the metric dimension of a graph*, Ars Combinatoria 2, pp. 191-195, 1976.

[11] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, *On a new class of codes for identifying vertices in graphs*, IEEE Transactions on Information theory, vol 44, pp. 599-611, 1998.

[12] J. Cáceres, C. Hernando, M. Mora, M. L. Puertas, I. M. Pelayo, C. Seara, *On the metric dimension of some families of graphs*, Electronic Notes in Discrete Mathematics, 22, pp. 129-133, 2005.

Dépôt légal : 2007 – 4ème trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications

Groupe des Ecoles des Télécommunications - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr

Département INFRES