



HAL
open science

Schéma de distance pour les pontés sans K_4

Victor Chepoi, Arnaud Labourel, Sébastien Ratel

► **To cite this version:**

Victor Chepoi, Arnaud Labourel, Sébastien Ratel. Schéma de distance pour les pontés sans K_4 . ALGOTEL 2020 – 22èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, Sep 2020, Lyon, France. ⟨hal-02872400⟩

HAL Id: hal-02872400

<https://hal.science/hal-02872400v1>

Submitted on 17 Jun 2020

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

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



HAL Authorization

Schéma de distance pour les pontés sans K_4

Victor Chepoi¹, Arnaud Labourel¹ et Sébastien Ratel¹

¹Aix Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France

Dans ce papier, nous décrivons un schéma de distance approchée pour la famille des graphes pontés sans K_4 , avec n sommets. Ce schéma utilise des étiquettes de $O(\log^3 n)$ bits. Étant données les étiquettes de deux sommets, la fonction de décodage du schéma calcule en temps constant une valeur entre la distance exacte séparant les sommets et quatre fois cette dernière.

Mots-clés : graphes bridged, schéma de distance

1 Introduction

Un schéma de distance (ou *DLS*) sur une famille de graphes \mathcal{G} consiste en une fonction d'encodage $C_G : V \rightarrow \{0, 1\}^*$ qui attribue une étiquette binaire à chaque sommet d'un graphe $G \in \mathcal{G}$, et en une fonction de décodage $D_G : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \mathbb{N}$ qui calcule la distance $d_G(u, v)$ dans G entre deux sommets u et v étant données leurs étiquettes, et sans connaissance supplémentaire sur le graphe. L'un des résultats principaux sur les DLS est le fait que la famille des graphes généraux sur n sommets admette un schéma de distance avec des étiquettes de $O(n)$ bits [GPPR04]. Ce schéma est asymptotiquement optimal puisqu'un simple argument de comptage de ces graphes suffit à montrer que $\Omega(n)$ bits sont nécessaires. Un autre résultat intéressant est qu'il existe un schéma de distance pour les arbres avec des étiquettes de $O(\log^2 n)$ bits. Assez récemment, Freedman et. al [FGNW17] ont obtenu un tel schéma permettant un décodage en temps constant. Plusieurs classes de graphes contenant les arbres possèdent aussi de tels schémas de distance avec étiquettes de $O(\log^2 n)$ bits. C'est le cas notamment des graphes de largeur arborescente bornée, des graphes distance-héréditaires, des graphes de largeur de clique ("clique-width") bornée ou encore des graphes de courbure négative ou nulle. Gavaille et al. [GPPR04] ont aussi donné une borne inférieure de $\Omega(\log^2 n)$ bits sur la taille des étiquettes pour les arbres, ce qui implique que tous les résultats mentionnés juste avant sont aussi optimaux. D'autres familles de graphes assez étudiées pour ce problème sont les graphes d'intervalles ou les graphes planaires (pour lesquels une borne inférieure de $\Omega(n^{1/3})$ bits est donnée dans [GPPR04], et une borne supérieure de $O(\sqrt{n})$ bits est fournie dans [GU16]). Il existe aussi des résultats concernant des schémas d'étiquetage de distance approchée, c'est-à-dire des schémas dont la fonction de décodage retourne la distance entre les sommets avec une erreur additive et/ou multiplicative. Pour les graphes arbitraires, le résultat le plus important, donné par Thorup & Zwick [TZ05], est un schéma de distance avec erreur multiplicative $(2k - 1)$ et tailles d'étiquettes $O(n^{1/k} \log^2 n)$ pour tout entier $k \geq 1$. Dans ce résumé, pour un $k \in \mathbb{N}^*$, nous appelons *schéma de distance k -approchée*, un schéma d'étiquetage de distance dans lequel la fonction de décodage retourne une valeur entre la distance exacte et k fois celle-ci.

Les graphes *pontés* ("bridged", en anglais) sont les graphes $G = (V, E)$ qui n'admettent pas de cycle isométrique de plus de trois sommets. Ils forment une sous-classe assez importante des graphes faiblement modulaires, une famille qui unifie un grand nombre de classes intéressantes de la théorie métrique des graphes par des caractérisations "local-vers-global". En effet, les graphes pontés sont exactement les graphes faiblement modulaires sans cycle induit de longueur 4 ou 5. Pour ce qui nous concerne ici, ils satisfont donc en particulier la *condition du triangle (CT)* suivante : Pour tout triplet de sommets $u, v, w \in V$ avec $k := d_G(u, v) = d_G(u, w)$ et $vw \in E$, il existe $x \in V$ tel que $d_G(u, x) = k - 1$ et $xv, xw \in E$.

Dans les graphes pontés, le voisinage des convexes est *convexe* (c.-à-d., contient tous les plus courts chemins entre chaque paire de ses sommets), ce qui est une propriété fondamentale des espaces CAT(0). Ils apparaissent aussi en théorie géométrique des groupes puisqu'ils coïncident avec le 1-squelette des complexes systoliques [Che00, JS06]. Nous présentons maintenant notre résultat principal.

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

2 L'idée du schéma

L'idée principale de notre schéma de distance pour les graphes pontés sans K_4 est similaire, dans les grandes lignes, à celle proposée dans [CLR19] pour les graphes médians sans cube. À savoir, le schéma est basé sur une partition récursive de nos graphes en une étoile et ses fibres : des cônes et des panneaux. En revanche, la structure et les propriétés de ces étoiles et de ces fibres sont très différentes de celles obtenues pour les graphes médians. Nous présentons dans cette section les propriétés principales qui permettent le fonctionnement de notre schéma pour les graphes pontés sans K_4 .

Soit $G = (V, E)$ un graphe ponté sans K_4 avec n sommets. L'algorithme d'encodage commence par rechercher un sommet médian m de G (c.-à-d., m minimise $x \mapsto \sum_{v \in V} d_G(x, v)$). Il calcule ensuite un voisinage à distance 2 particulier de m , que l'on appelle l'étoile $\text{St}(m)$ de m (et que nous définissons dans la section suivante). Chaque sommet de G peut être associé à un unique sommet de $\text{St}(m)$. Cela nous permet de définir des fibres : pour un $x \in \text{St}(m)$, la fibre $F(x)$ correspond à l'ensemble des sommets de G associés à x . Nous pouvons alors montrer que l'ensemble des fibres de $\text{St}(m)$ constitue une partition de G . De plus, le choix de m en tant que sommet médian nous permet d'affirmer que chaque fibre contient au plus la moitié des sommets. Les fibres $F(x)$ de $\text{St}(m)$ ne sont pas convexes. Elles sont, en revanche, connexes et isométriques, et induisent, par conséquent, des graphes pontés. Cela nous permet de re-partitionner récursivement, sans accumuler d'erreur sur les distances à chaque étape. Enfin, nous devons étudier les frontières et la bordure de ces fibres (c.-à-d., respectivement, l'ensemble des sommets d'une fibre ayant un voisin dans une autre fibre en particulier, et l'union de toutes les frontières d'une fibre). Ces bordures ne sont pas tout à fait des arbres, mais des "arbres étoilés". Malheureusement, ces arbres étoilés ne sont pas non plus isométriques. Cela explique, en partie, que nous obtenions un schéma de distance approchée. En effet, les distances dans les bordures peuvent être le double des distances du graphe. Nous distinguerons deux types de fibres $F(x)$ selon la distance entre x et m : les panneaux pour une distance 1 ; les cônes pour une distance 2. L'un des résultats principaux pour obtenir notre schéma compact est qu'un sommet dans un panneau admet deux "sorties" sur la bordure de ce panneau, et qu'un sommet dans un cône admet une "entrée" par frontière du cône (et il n'y en a que deux). Ces "entrées" ou "sorties" d'un sommet u d'une fibre $F(x)$ auront la garantie d'être sur un chemin de longueur inférieure à quatre fois celle d'un plus court chemin entre u et v , pour n'importe quel sommet $v \notin F(x)$. Ainsi, à chaque étape, chaque sommet n'aura à stocker des informations relatives qu'à deux autres sommets. Sans cette propriété, notre schéma utiliserait des étiquettes de taille linéaire puisque, un panneau pouvant être voisin d'un nombre linéaire de fibres, un sommet dans un panneau pourrait avoir besoin de stocker des informations relatives à un nombre linéaire de frontières pour garantir que toutes les distances puissent être calculée avec une erreur (multiplicative) constante. En autorisant une erreur multiplicative de 4, nous verrons que dans presque tous les cas, la distance retournée entre deux sommets u et v sera celle passant par le centre m d'une étoile $\text{St}(m)$ de la partition à une certaine étape récursive, à déterminer. Les lemmes 1 et 4 indiquent respectivement quand cette valeur correspond à une distance exacte et quand elle donne une distance approchée (à facteur 2). Les cas dans lesquels u et v sont dans des fibres "trop proches" sont ceux nécessitant un calcul plus technique et conduisant à une erreur multiplicative de 4 dans le pire des cas.

3 Étoiles et fibres

Soit $m \in V$ un sommet médian d'un graphe ponté $G = (V, E)$ sans K_4 . Pour tout sommet $u \in V$, nous pouvons montrer que l'ensemble $\text{Pr}(u, N[m])$ des sommets du voisinage fermé $N[m]$ de m à distance minimale de u ($\text{Pr}(u, N[m])$ est appelé *projection de u sur $N[m]$*) consiste soit en un unique sommet, soit en deux sommets reliés par une arête. Si $u \in V$ est un sommet avec deux projections x et y sur $N[m]$, alors par (CT), il existe un sommet u' à distance $d_G(x, u) - 1$ de u , et relié à x et à y . L'étoile $\text{St}(m)$ d'un sommet $m \in V$ correspond à son voisinage $N[m]$ ainsi qu'à tous les triangles que l'on peut dériver (par (CT)) avec les

Schéma de distance pour les pontés sans K_4

projections sur $N[m]$ étant des arêtes. Notons que cette étoile peut avoir une structure complexe puisque les voisinages ouverts des sommets des graphes pontés sans K_4 sont en bijection avec l'ensemble des graphes de maille au moins 6. Nous pouvons en déduire que les graphes pontés sans K_4 peuvent contenir tous les graphes comme mineur.

Soit x un sommet de $\text{St}(m)$. Si $x \in N[m]$, nous définissons la *fibre* $F(x)$ de x respectivement à $\text{St}(m)$ comme l'ensemble de tous les sommets de G ayant x pour unique projection sur $N[m]$. Sinon (c.-à-d., si $d_G(x, m) = 2$), $F(x)$ dénote l'ensemble des sommets ayant une arête pour projection sur $N[m]$ et qui sont plus proche de x que des extrémités de cette arête. Nous obtenons alors que $\mathcal{F}_m := \{F(x) : x \in \text{St}(m)\}$ est une partition de G dont les parties induisent des graphes pontés isométriques d'au plus $|V|/2$ sommets.

Une fibre $F(x)$ est appelée un *panneau* si $d_G(x, m) = 1$, ou un *cône* si $d_G(x, m) = 2$. Deux fibres $F(x)$ et $F(y)$ sont k -voisines si $d_{\text{St}(m) \setminus \{m\}}(x, y) = k$. La *frontière* $\partial_{F(y)} F(x)$ de $F(x)$ vers $F(y)$ est l'ensemble $\{u \in F(x) : \exists v \in F(y) \text{ t.q. } uv \in E\}$ des sommets de $F(x)$ ayant un voisin dans $F(y)$. La *bordure* $\partial^* F(x)$ de $F(x)$ est l'union de ses frontières. Nous pouvons montrer que, pour tout sommet u de la bordure $\partial^* F(x)$ d'une fibre $F(x)$, il existe un unique plus court chemin entre u et x inclus dans $\partial^* F(x)$. L'union des plus courts chemins dans $\partial^* F(x)$ de x à tous les $u \in \partial^* F(x)$ forme un arbre que nous appelons *arbre étoilé*. Cet arbre n'est pas forcément isométrique, ni même induit, mais la distance entre toute paire de sommets (u, v) de cet arbre étoilé est au plus doublée par rapport à la distance dans G . Enfin, l'un de nos résultats principaux est que, pour tout sommet $u \in V$, il existe deux sommets u_1 et u_2 dans l'arbre T induit par $\partial^* F(x)$ (que nous appelons *sorties de u sur T* , et dont la description ne sera pas donnée par souci de concision) tels que, pour tout $v \in V(T)$, $\min\{d_G(u, u_1) + d_T(u_1, v), d_G(u, u_2) + d_T(u_2, v)\} \leq 2 \cdot d_G(u, v)$.

4 Classification des paires de sommets et plus courts chemins

Nous présentons dans cette section les cas dans lesquels le centre m de l'étoile $\text{St}(m)$ appartient à l'ensemble $I(u, v)$ des sommets sur un plus court chemin entre deux sommets u et v (lemme 1), ceux dans lesquels passer par ce sommet m peut conduire à une erreur multiplicative d'au plus 2 (lemme 4), et dans quels cas nous pourrions faire une erreur multiplicative d'au plus 4 (lemmes 2 et 3). Ces cas sont cruciaux pour le schéma de distance présenté dans la section suivante.

Dans le reste de cette section, $m \in V$ dénote un sommet médian de G . Nous posons aussi, $x, y \in \text{St}(m)$ et $(u, v) \in F(x) \times F(y)$. Si $F(x) = F(y)$, alors u et v sont *proches*. Nous pouvons remarquer qu'à partir d'une certaine étape récursive, u et v ne peuvent plus être proches. Ce cas n'est donc pas vraiment à traiter ici. Lorsque $m \in I(u, v)$, u et v sont *séparés*.

Lemme 1. *u et v sont séparés (respectivement à m) si et seulement si $F(x) \neq F(y)$ sont : (i) deux panneaux k -voisins, où $k \geq 2$; (ii) un panneau et un cône k -voisins, où $k \geq 3$; (iii) ou deux cônes k -voisins, où $k \geq 4$.*

Quand $F(x)$ et $F(y)$ sont 1-voisines, l'une d'elles est un panneau et l'autre un cône, u et v sont *1pc-voisins*.

Lemme 2. *Si u et v sont 1pc-voisins, considérons que $F(x)$ est le panneau et notons T l'arbre étoilé induit par $\partial^* F(x)$. Soient u_1 et u_2 les sorties de u sur T , et soit $v' \in \text{Pr}(v, T)$ l'unique sommet de T à distance minimale de x (appelé entrée de v sur T). Alors*

$$d_G(u, v) \leq \min\{d_G(u, u_1) + d_T(u_1, v'), d_G(u, u_2) + d_T(u_2, v')\} + d_G(v', v) \leq 4 \cdot d_G(u, v).$$

Si $F(x)$ et $F(y)$ sont deux cônes 2-voisins, alors u et v sont *2cc-voisins*.

Lemme 3. *Si u et v sont 2cc-voisins, considérons le panneau $F(w)$ 1-voisin de $F(x)$ et $F(y)$, et posons $T := \partial^* F(w)$. Soient u' et v' les entrées respectives de u et v sur T . Alors*

$$d_G(u, v) \leq d_G(u, u') + d_T(u', v') + d_G(v', v) \leq 4 \cdot d_G(u, v).$$

Dans les cas restant, u et v sont *quasi-séparés*.

Lemme 4. *Si u et v sont quasi-séparés, alors $d_G(u, m) + d_G(m, v) \leq 2 \cdot d(u, v)$.*

5 Schéma de distance 4-approchée

Nous décrivons maintenant, sans entrer dans les détails, un schéma d'étiquetage de distance 4-approchée pour les graphes pontés sans K_4 , basé sur les propriétés établies dans les sections précédentes. La description précise du schéma et les détails des preuves sont donnés dans la thèse [Rat19, Chapitre 8].

5.1 Encodage

Soit u un sommet de G . Nous décrivons ici la partie d'étiquette $L_i(u)$ de u construite, à l'étape (réursive) i , par la procédure d'encodage. $L_i(u)$ est constituée de trois parties : “St”, “1st” et “2nd”. La première partie, L_i^{St} , contient les informations relatives à l'étoile $\text{St}(m)$ du médian m de l'étape i correspondante : l'identifiant unique de m dans G ; la distance $d_G(u, m)$ de u à m ; et un étiquetage particulier de la fibre de u sur $\text{St}(m)$. Ce dernier identifiant permet de déterminer à quel type de fibre appartient u , ainsi que le statut (proches, séparés, 1pc-voisins ou autre) de la paire (u, v) pour tout autre sommet $v \in V$. Les deux dernières parties, L_i^{1st} et L_i^{2nd} , contiennent les informations relatives aux deux entrées (ou sorties) u_1 et u_2 de u sur : les bordures des deux panneaux 1-voisins $F(w_1)$ et $F(w_2)$ de $F(x)$, si $F(x)$ est un cône; ou sur la bordure de $F(x)$ si $F(x)$ est un panneau. Si $F(x)$ est un cône, alors L_i^{1st} contient une étiquette de distance exacte de u_1 dans l'arbre étoilé $\partial^*F(w_1)$, ainsi que la distance $d_G(u, u_1)$ entre u et u_1 . Et si $F(x)$ est un panneau, alors L_i^{1st} contient une étiquette de distance exacte de u_1 dans la bordure de $F(x)$ (celui décrit dans [FGNW17], par exemple), ainsi que la distance $d_G(u, u_1)$ entre u et u_1 . La partie L_i^{2nd} est d'une forme identique à L_i^{1st} , mais où u_2 remplace u_1 (et où $\partial^*F(w_2)$ remplace $\partial^*F(w_1)$ si $F(x)$ est un cône).

Puisqu'à chaque étape réursive, le nombre de sommets de chaque partie est au moins divisé par deux, le nombre d'étapes sera de l'ordre de $O(\log |V|)$. À chaque étape réursive, les sommets ajoutent à leur étiquettes un nombre constant d'informations, la plus coûteuse d'entre elles étant une étiquette de distance dans un arbre nécessitant $O(\log^2 |V|)$ bits. Ainsi, notre schéma utilise $O(\log^3 |V|)$ bits par sommet.

5.2 Requêtes de distances

Étant données les étiquettes $L(u)$ et $L(v)$ de deux sommets, la fonction de décodage commence par déterminer le statut de la paire (u, v) . Pour cela, elle recherche le médian séparateur m de u et v , ou plutôt, la partie i d'étiquette correspondant à l'étape ayant fait émerger ce médian. Notons que la même astuce que celle présentée dans [CLR19] nous permet de trouver ce médian en temps constant. Une fois cette étape effectuée, la fonction de décodage déterminera que les deux sommets sont 1pc-voisins si et seulement l'identifiant de la fibre de l'un est inclus dans celui de la fibre de l'autre. Dans ce cas, la fonction de décodage fera appel à une procédure basée sur le lemme 2. C'est-à-dire que, cette procédure retournera $\min\{d_G(u, u_1) + d_T(u_1, v'), d_G(u, u_2) + d_T(u_2, v')\} + d_G(v', v)$, où nous avons supposé que u est dans un panneau (et v dans un cône), où u_1, u_2 et v' sont contenus dans les parties d'étiquettes $L_i^{\text{2nd}}(u)$ (ou $L_i^{\text{1st}}(u)$) et $L_i^{\text{2nd}}(v)$ (ou $L_i^{\text{1st}}(v)$), et où les distances $d_T(u_1, v')$ et $d_T(u_2, v')$ sont obtenues en décodant les étiquettes de distance d'arbre de u_1, u_2 et v' dans T , aussi contenus dans ces parties d'étiquettes. Les sommets u et v seront déterminés 2cc-voisins si les identifiants de leurs fibres s'intersectent sur un singleton. Dans ce cas, le décodage se fera via une procédure basée sur le Lemme 3. Dans tous les cas restant (c.-à-d., u et v séparés ou quasi-séparés), l'algorithme de décodage retournera la somme $d_G(u, m) + d_G(v, m)$. Cette somme étant comprise entre $d_G(u, v)$ et $2 \cdot d_G(u, v)$ d'après les lemmes 1 et 4.

Références

- [Che00] V. Chepoi. Graphs of some CAT(0) complexes. *Adv. Appl. Math.*, 24 :125–179, 2000.
- [CLR19] V. Chepoi, A. Labourel, and S. Ratel. Distance labeling schemes for cube-free median graphs. In *MFCS*, volume 138, pages 15 :1–15 :14, 2019.
- [FGNW17] O. Freedman, P. Gawrychowski, P. K. Nicholson, and O. Weimann. Optimal distance labeling schemes for trees. In *PODC*, pages 185–194. ACM, 2017.
- [GPPR04] C. Gavaille, D. Peleg, S. Pérennès, and R. Raz. Distance labeling in graphs. *J. Algorithms*, 53 :85–112, 2004.
- [GU16] P. Gawrychowski and P. Uznanski. A note on distance labeling in planar graphs. *CoRR*, abs/1611.06529, 2016.
- [JS06] T. Januszkiewicz and J. Świątkowski. Simplicial nonpositive curvature. *Publ. IHES*, 104 :1–85, 2006.
- [Rat19] S. Ratel. *Densité, VC-dimension et étiquetages de graphes*. PhD thesis, 2019.
- [TZ05] M. Thorup and U. Zwick. Approximate distance oracles. *Journal of ACM*, 52 :1–24, 2005.