



HAL
open science

Algorithme exact et approché pour le calcul de l'hyperbolicité d'un graphe

Nathann Cohen, David Coudert, Aurélien Lancin

► **To cite this version:**

Nathann Cohen, David Coudert, Aurélien Lancin. Algorithme exact et approché pour le calcul de l'hyperbolicité d'un graphe. AlgoTel 2013 - 15èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2013, Pornic, France. pp.1-4. hal-00818441

HAL Id: hal-00818441

<https://hal.science/hal-00818441>

Submitted on 29 Apr 2013

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.

Algorithme exact et approché pour le calcul de l'hyperbolicité d'un graphe[†]

Nathann Cohen¹, David Coudert² et Aurélien Lancin²

¹LRI, Université Paris-Sud 11, France

²COATI, INRIA, I3S, CNRS, Univ. Nice Sophia, France.

Nous proposons un algorithme simple et efficace pour calculer l'hyperbolicité de grands graphes dont la complexité temporelle est fonction de la distribution des plus courts chemins dans les composantes biconnexes du graphe et de la valeur de l'hyperbolicité. Nous montrons également comment réduire la taille de l'instance en utilisant une décomposition par des cliques-séparatrices. L'algorithme peut de plus être utilisé pour fournir une approximation de l'hyperbolicité à un facteur multiplicatif ou une constante additive donnés. Nous évaluons les performances de notre algorithme sur des cartes des systèmes autonomes de l'Internet (CAIDA et DIMES).

Keywords: Hyperbolicité, algorithme, graphe, approximation

1 Introduction

De nombreuses propriétés structurelles de grands graphes, en particulier de l'Internet, ont été étudiées ces dernières années en vue de mieux comprendre leur structure, de les modéliser et de proposer de nouveaux algorithmes de routage. Dans cet article, nous nous intéressons à la notion d'*hyperbolicité*, introduite par Gromov [Gro87], qui indique la différence entre la mesure des distances dans le graphe et dans un arbre. Par exemple, lorsque la topologie de l'Internet est plongée dans un espace hyperbolique, il a été montré qu'un algorithme de routage glouton utilisant les coordonnées des sommets et les distances entre sommets dans cet espace métrique a une faible élongation (*stretch*) par rapport aux plus courts chemins [BPK10, PKBV09]. Cette notion fournit aussi des informations sur la dispersion des plus courts chemins (distance entre les sommets de deux plus courts chemins de s à t). C'est donc une notion importante pour la conception d'algorithmes de routage.

Une définition de l'hyperbolicité δ d'un graphe, appelée *condition des 4 points*, est la suivante : Soit G un graphe connexe et $d(a, b)$ la distance entre les sommets a et b de G .

Définition 1. *Un graphe simple sans boucle G est dit δ -hyperbolique si pour tout quadruplet $a, b, c, d \in V$, les deux plus grandes des sommes $S_1 = d(a, b) + d(c, d)$, $S_2 = d(a, c) + d(b, d)$, et $S_3 = d(a, d) + d(b, c)$ diffèrent d'au plus 2δ . L'hyperbolicité de G désigne la valeur du plus petit δ vérifiant cette définition.*

L'arbre et la clique sont 0-hyperbolique, le cycle d'ordre n est $(n/4)$ -hyperbolique, et la grille $n \times m$ avec $n \leq m$ est $n - 1$ -hyperbolique. Ceci indique bien que les plus courts chemins entre deux sommets de la grille peuvent-être très éloignés. Dans cet article, nous cherchons à déterminer la valeur exacte de l'hyperbolicité du graphe des systèmes autonomes (AS) de l'Internet qui avait été estimée avec des heuristiques [dMSV11]. Le meilleur algorithme exact a une complexité en $O(n^{3.69})$ [FIV12], ce qui est prohibitif sur les graphes considérés ($O(10^{16})$ opérations pour un graphe biconnexe à 25 000 sommets, soit des semaines de calculs).

Dans cet article, nous proposons un nouvel algorithme exact combinant des méthodes de décomposition (section 2) et un ordre de parcours des données permettant de couper l'espace de recherche (section 3). Si cet algorithme a une complexité théorique en $O(n^4)$, il est très efficace en pratique (section 4) et nous a permis de déterminer l'hyperbolicité de toutes les cartes des AS collectées par CAIDA en un temps raisonnable. De plus, cet algorithme peut-être utilisé comme un algorithme d'approximation avec un facteur multiplicatif ou une constante additive donnés.

[†]Ce travail a été effectué dans le cadre du projet européen FP7 EULER et des projets ANR AGAPE et GRATEL.

2 Méthodes de décomposition

Dans cette section, nous expliquons comment utiliser une décomposition du graphe $G = (V, E)$ par des cliques-séparatrices [Tar85] pour faciliter le calcul de l'hyperbolicité. Nous considérons que G est biconnexes et d'hyperbolicité $\delta > 1$. En effet, l'hyperbolicité d'un graphe est la valeur maximale de l'hyperbolicité de ses composantes biconnexes (par l'inégalité triangulaire), et les graphes d'hyperbolicité $\delta \leq 1$ peuvent être reconnus en temps linéaire [BKM01, B03, KM02].

Nous appelons $X \subset V$ une $(A|B)$ -clique-séparatrice si $G[X]$ est une clique séparant $A \setminus X$ de $B \setminus X$, et un *atome* est un sous-graphe maximal de G ne contenant pas de clique-séparatrice. Une décomposition de G par des cliques-séparatrices est donnée par l'ensemble de ses *atomes* et se calcule en temps $O(|V| \cdot |E|)$ [Tar85].

En utilisant l'inégalité triangulaire, nous montrons que si $a_1, a_2 \in A$ et $b_1, b_2 \in B$ alors $\delta(a_1, a_2, b_1, b_2) \leq 1$. Ainsi, nous pouvons ignorer de tels quadruplets. Il restent à considérer i) tous les quadruplets appartenant à un même atome, ii) les quadruplets avec $a \in A$ séparé de $b_1, b_2, b_3 \in B$ par une $(A|B)$ -clique-séparatrice X .

Pour le cas ii), nous commençons par choisir $x_i \in X$ tel que $d(a, b_i) = d(a, x_i) + d(x_i, b_i)$ et poser $m = d(a, X)$. Alors, si nous ajoutons un sommet v_a à G si et seulement si $d(a, x_i) = m$, nous obtenons $\delta(a, b_1, b_2, b_3) = \delta(v_a, b_1, b_2, b_3)$. De plus, si $x_1 = x_2 = x_3$ nous avons $\delta(a, b_1, b_2, b_3) = \delta(x_1, b_1, b_2, b_3)$. Ensuite, nous construisons le *substitut* A^* de $A \setminus X$ qui contient un sommet $v_{i,j} \in A^*$ adjacent à $x_i, x_j \in X$ si il existe un sommet $a \in A \setminus X$ tel que $d(a, x_i) = d(a, x_j) = d(a, X)$ et $x_i \neq x_j$. Nous avons $|A^*| \leq \binom{|X|}{2}$. Clairement, si $|A^*| > |A|$ alors nous utilisons directement A . Les sommets de A tels que $d(a, x_1) = d(a, x_2) = d(a, x_3)$ sont émülés par les sommets $v_{1,2}, v_{1,3}, v_{2,3} \in A^*$. Ensuite, si $(V \setminus A)^*$ est l'union des substituts des composantes connexes de $V \setminus A$, nous obtenons :

Théorème 2. *Soit $G = (V, E)$ un graphe et A_i l'ensemble des atomes de la décomposition par des cliques-séparatrices. Nous avons $\delta(G) = \max\{1, \max_i \delta(G[A_i \cup (V \setminus A_i)^*])\}$.*

Clairement, nous avons $\delta(G[A_i \cup (V \setminus A_i)^*]) \geq \delta(G[A_i])$. Par ailleurs, si $x_a \in X_a$ est un voisin de $a \in A^*$ et $b, c, d \notin A^*$ nous obtenons par l'inégalité triangulaire que $\delta(a, b, c, d) \leq \delta(x_a, b, c, d) + 1/2$. Si de plus d est séparé de a, b, c, x_a par une clique X_d alors nous pouvons montrer que $\delta(a, b, c, d) \leq \delta(x_a, b, c, x_d) + 1/2$. En appliquant au plus deux fois ce résultat à un quadruplet, nous montrons que l'hyperbolicité des atomes sans les substituts donne une approximation de l'hyperbolicité de G à un près.

Théorème 3. *Soit $G = (V, E)$ un graphe et A_i l'ensemble des atomes de la décomposition par des cliques-séparatrices. Nous avons $\max_i \delta(G[A_i]) \leq \delta(G) \leq \max_i \delta(G[A_i]) + 1$.*

Ainsi, quand $\delta(G) > 1$, nous décomposons le graphe en atomes par des cliques-séparatrices, nous ajoutons à chaque atome le substitut de $V \setminus A$ et nous appliquons sur chacun des graphes résultants l'algorithme 1.

3 Algorithme pour le calcul de l'hyperbolicité

L'idée de l'algorithme est de parcourir les quadruplets en commençant par ceux pouvant fournir la plus grande valeur d'hyperbolicité. Plus précisément, nous parcourons les quadruplets de façon à ce que le quadruplet a, b, c, d soit testé avant le quadruplet a', b', c', d' si $\min(d(a, b), d(c, d)) > \min(d(a', b'), d(c', d'))$. Ceci repose sur le lemme 4 qui montre en outre que $\delta \leq D/2$, où D est le diamètre du graphe.

Lemme 4. *Soient $G = (V, E)$ un graphe connexe, $a, b, c, d \in V$, $S_1 = d(a, b) + d(c, d)$, $S_2 = d(a, c) + d(b, d)$, et $S_3 = d(a, d) + d(b, c)$. Si $S_1 \geq \max\{S_2, S_3\}$, alors $\delta(a, b, c, d) \leq \min\{d(a, b), d(c, d)\} / 2$.*

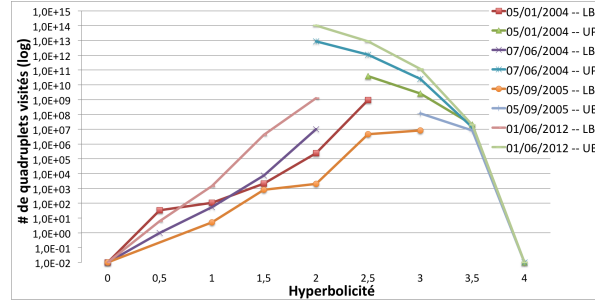
Démonstration. $S_2 + S_3 = (d(a, c) + d(b, d)) + (d(a, d) + d(b, c)) = (d(a, c) + d(b, c)) + (d(a, d) + d(b, d))$. L'inégalité triangulaire permet de déduire $S_2 + S_3 \geq 2 \cdot d(a, b)$. Comme $S_1 \geq \max\{S_2, S_3\}$, nous avons $2 \cdot \delta(a, b, c, d) = S_1 - \max\{S_2, S_3\} \leq S_1 - (S_2 + S_3)/2 \leq S_1 - d(a, b) = d(c, d)$. De la même manière $2 \cdot \delta(a, b, c, d) \leq d(a, b)$. \square

Grace au lemme 4, à chaque étape de l'algorithme, $d(c, d)$ est une borne supérieure de la valeur de $\delta^*(a, b, c, d)$. Si la borne inférieure courante est δ^* , aucun quadruplet tel que $d(c, d) \leq \delta^*$ ne peut être utilisé pour améliorer la borne inférieure. L'exploration des quadruplets est ainsi coupée. Comme notre algorithme peut avoir à parcourir les $\binom{n}{2}$ paires de sommets (notamment lorsque G est chordal), sa complexité dans le

Algorithm 1: Hyperbolicité

Require: $G = (V, E)$ est un graphe biconnexe.
Ensure: δ , l'hyperbolicité de G (avec $\delta = 2\delta^*$).
1: Soit `paires` la liste des $\binom{n}{2}$ paires de sommets triée par distances décroissantes
2: $\delta^* := 0$
3: **for** $0 \leq i < \binom{n}{2}$ **do**
4: $(c, d) := \text{paires}[i]$
5: **if** $d(c, d) \leq \delta^*$ **then**
6: **break**
7: **for** $0 \leq j < i$ **do**
8: $(a, b) := \text{paires}[j]$
9: $\delta^* := \max\{\delta^*, \delta^*(a, b, c, d)\}$
10: **if** $d(a, b) \leq \delta^*$ **then**
11: **break**
12: **return** $\delta^*/2$

FIGURE 1: Nombre de quadruplets testés pour atteindre des bornes inférieures (LB) et supérieures (UB) sur l'hyperbolicité de cartes CAIDA. Les bornes inférieures évoluent de gauche à droite, et de droite à gauche pour les bornes supérieures.



pire des cas est de $O(n^4)$. L'exécution de notre algorithme requiert le calcul de la matrice des distances et le tri des paires de sommets par ordre de distances décroissantes (ceci revient à associer à chaque entier ℓ la liste des paires de sommets à distance ℓ). Ces opérations ont respectivement une complexité en $O(n(n+m))$ et $O(n^2)$ pour des graphes à n sommets et m arêtes.

La complexité théorique de l'algorithme peut être paramétrée par la valeur optimale de l'hyperbolicité et la distribution des distances. Notons P_ℓ le nombre de paires (a, b) telles que $d(a, b) = \ell$. L'algorithme ne considère que les paires de paires $((a, b), (c, d))$ telles que $D \geq d(c, d) = \ell_2 \geq d(a, b) = \ell_1 \geq 2\delta$. En particulier, pour les grilles $n \times n$ où l'hyperbolicité est donnée par le quadruplet $(0, 0), (0, n-1), (n-1, 0), (n-1, n-1)$, l'algorithme termine après avoir testé un seul quadruplet.

Proposition 5. *Etant donné un graphe δ -hyperbolique de diamètre D avec P_ℓ paires de sommets à distance ℓ , la complexité des lignes 2-12 de l'algorithme est en $O\left(\left(\sum_{\ell=2\delta}^D P_\ell\right)^2\right)$.*

A chaque étape de l'algorithme, $d(c, d)/2$ est une borne supérieure et $\delta^*/2$ une borne inférieure. Nous pouvons donc obtenir une approximation du résultat à un facteur multiplicatif ou à une constante additive en stoppant l'exécution de l'algorithme lorsque l'approximation demandée est prouvée.

4 Résultats expérimentaux

Notre algorithme est maintenant disponible dans Sage [S⁺13] et Grph[‡]. Nous avons utilisé l'implémentation de Sage pour calculer l'hyperbolicité de l'ensemble des 190 cartes CAIDA des AS de 2004 à 2012 (Fig. 2). Les temps de calculs vont de quelques secondes pour les cartes d'hyperbolicité 3 à 8 jours pour la carte de 2012 d'hyperbolicité 2. Nous observons une augmentation du temps de calcul avec la taille du graphe et la diminution de la valeur de l'hyperbolicité. Une grande partie du temps est passée à prouver l'optimalité de la solution (Fig. 1). Les résultats complets avec certificats (quadruplets tels que $\delta(a, b, c, d) = \delta(G)$ pour chaque carte) sont disponibles dans [CCL12]. Les méthodes de décompositions de la section 2 réduisent significativement la taille des graphes à considérer. La carte CAIDA du 01/06/2012 a 41203 sommets, sa composante biconnexe principale 25815, le plus grand atome donné par la décomposition par des cliques-séparatrices a 18717 sommets et avec le substitut nous obtenons un graphe à 19241 sommets.

En dehors de l'observation que l'hyperbolicité des cartes des AS de l'Internet a globalement diminuée au cours du temps de 2.5 à 2, nous avons identifié la présence d'un grand nombre de sommets simpliciaux (de l'ordre de 16% des sommets), principalement des triangles. Ceci donne des informations supplémentaires pour la modélisation de ces graphes et la conception d'algorithmes de routage.

‡. <http://grph.inria.fr>

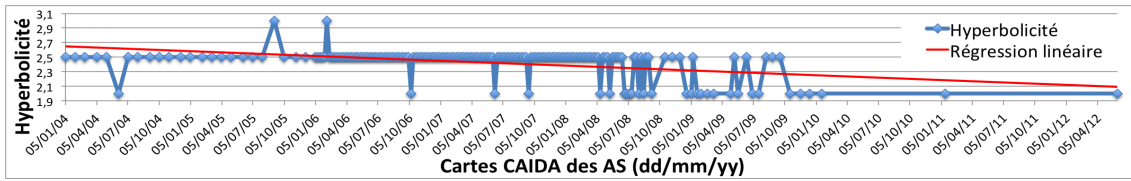


FIGURE 2: Evolution de l'hyperbolicité pour les cartes des AS de CAIDA depuis 2004.

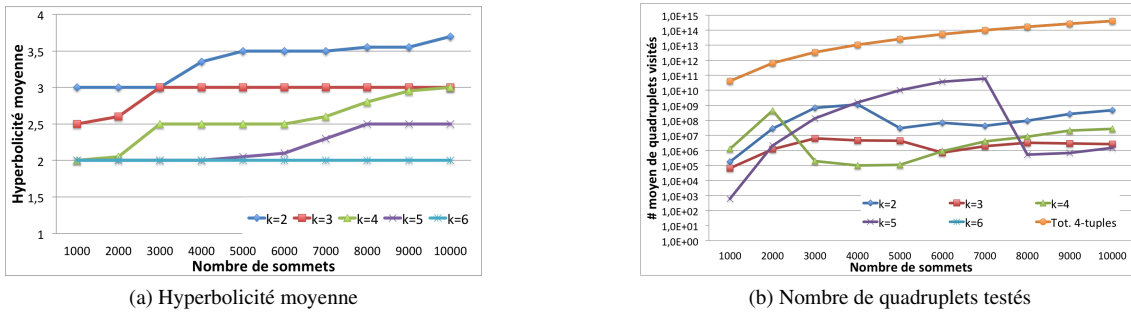


FIGURE 3: Barabasi-Albert.

La figure 3 montre l'hyperbolicité moyenne (sur 100 graphes) et le nombre de quadruplets testés par l'algorithme pour des graphes aléatoires du modèle Barabasi-Albert. Ces graphes sont construits par attachement préférentiel, chaque nouveau sommet étant connecté à k sommets existants avec une probabilité proportionnelle au degré de ces sommets. Les résultats montrent que notre algorithme ne visite qu'une petite partie des quadruplets (moins de 0.1%). Les irrégularités dans les courbes s'expliquent par l'évolution de la distribution des distances corrélée à la valeur de l'hyperbolicité. Il nous reste à établir analytiquement le temps d'exécution de l'algorithme sur ces graphes.

Références

- [B03] H.-J. Bandelt and V. . 1-hyperbolic graphs. *SIAM J. Discrete Math.*, 16(2) :323–334, 2003.
- [BKM01] G. Brinkmann, J. H. Koolen, and V. Moulton. On the hyperbolicity of chordal graphs. *Annals of Combinatorics*, 2001.
- [BPK10] M. Boguna, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with Hyperbolic Mapping. *Nature Communications*, 1(62), Oct 2010.
- [CCL12] N. Cohen, D. Coudert, and A. Lancin. Exact and approximate algorithms for computing the hyperbolicity of large-scale graphs. Rapport de recherche RR-8074, INRIA, September 2012.
- [dMSV11] F. de Montgolfier, M. Soto, and L. Viennot. Treewidth and hyperbolicity of the internet. In *10th IEEE International Symposium on Network Computing and Applications (IEEE NCA)*, pages 25–32, Boston, 2011.
- [FIV12] H. Fournier, A. Ismail, and A. Vigneron. Computing the Gromov hyperbolicity of a discrete metric space. Technical Report arXiv :1210.3323, ArXiv, October 2012.
- [Gro87] M. Gromov. Hyperbolic groups. *Essays in Group Theory*, 8 :75–263, 1987.
- [KM02] J. H. Koolen and V. Moulton. Hyperbolic bridged graphs. *Eur. J. Comb.*, 23(6) :683–699, 2002.
- [PKBV09] F. Papadopoulos, D. Krioukov, M. Boguna, and A. Vahdat. Greedy forwarding in scale-free networks embedded in hyperbolic metric spaces. *SIGMETRICS Performance Evaluation Review*, 37(2) :15–17, 2009.
- [S⁺13] W. A. Stein et al. *Sage Mathematics Software (Version 5.6)*. The Sage Development Team, 2013. <http://www.sagemath.org>.
- [Tar85] R. E. Tarjan. Decomposition by clique separators. *Discrete Mathematics*, 55(2) :221 – 232, 1985.