



**HAL**  
open science

## Decomposing a Graph into Shortest Paths with Bounded Eccentricity

Etienne E. Birmelé, Fabien de Montgolfier, Léo Planche, Laurent Viennot

► **To cite this version:**

Etienne E. Birmelé, Fabien de Montgolfier, Léo Planche, Laurent Viennot. Decomposing a Graph into Shortest Paths with Bounded Eccentricity. 2017. hal-01511357v1

**HAL Id: hal-01511357**

**<https://hal.science/hal-01511357v1>**

Preprint submitted on 20 Apr 2017 (v1), last revised 14 Sep 2017 (v2)

**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.

# Decomposing a Graph into Shortest Paths with Bounded Eccentricity

Etienne Birmelé<sup>1</sup>, Fabien de Montgolfier<sup>1\*</sup>,  
Léo Planche<sup>12</sup>, and Laurent Viennot<sup>32\*</sup>

<sup>1</sup> MAP5, UMR CNRS 8145, Univ. Sorbonne Paris Cité

<sup>2</sup> Université Paris Diderot - Paris 7

<sup>3</sup> Inria

**Abstract.** We introduce the problem of hub-laminar decomposition which generalizes that of computing a shortest path with minimum eccentricity. It consists in decomposing a graph into several paths that collectively have small eccentricity and meet only near their extremities. The problem is also related to that of binning appearing in biology in the context of metagenomics. We show that a graph having such a decomposition with sufficient long paths can be decomposed with approximated guarantees on the parameters of the decomposition.

## 1 Introduction

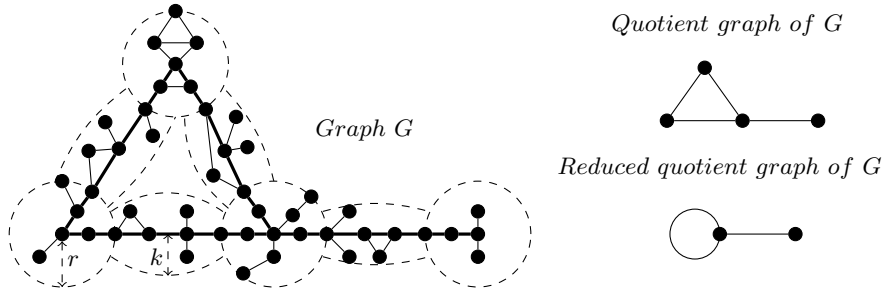
The goal of this paper is to extend the MESP (Minimum Eccentricity Shortest Path) Problem from Dragan and Leiter [4] and the related problem of recognizing  $k$ -laminar graphs from Völkel *et al.* [11]. Both consist in finding a shortest path (in the sense that no path joining the same endpoints is shorter)  $k$ -covering a graph (every vertex is at distance at most  $k$  from that path). The  $k$ -laminar problem additionally requires that path to be a diameter. Relationships between the two parameters are derived in [3].

To generalize this problem to several paths, we introduce the problem of decomposing a graph into subgraphs with bounded shortest-path eccentricity. More precisely, we introduce the hub-laminar decomposition as a set of paths that  $k$  cover the graph and meet only at their extremities. To formalize this property, we introduce the notion of hub, that is a ball with fixed radius  $r$  centered at a path endpoint. The laminar associated to a path is the set of nodes  $k$ -covered by the path. Our definition requires that an edge between two nodes belonging to two different laminars must appear in a hub. (See Fig. 1 for an example.) The main result of the paper is that computing such decomposition becomes tractable when hub centers are far enough one from another. The MESP problem is equivalent to hub-laminar decomposition with one laminar. So is the problem of recognizing a  $k$ -laminar graph with the additional requirement that hub centers are  $D$  far apart where  $D$  is the diameter of the graph.

The motivation for such generalization is twofold. First, the MESP problem was introduced because of its relationship with that of embedding a graph into

---

\* Supported by IRIF (CNRS UMR 8243) and Inria project-team GANG.



**Fig. 1.** Illustration of an hub-laminar decomposition with  $r = 2, k = 1, l = 6, \lambda = 4$ . Every vertex is at distance  $r$  from a hub center (vertices at the center of dashed circles) or at distance  $k$  from a laminar path (paths with bold edges between hub centers).

the line with constant distortion [2]. Such an embedding allows to represent the distances between nodes with a single number per node (its position in the line). An approximation of the distance between two nodes is then obtained as the difference between the two positions. A similar motivation is that of representing the distances in a graph with a succinct data-structure and with small distortion. We show that a graph having a hub-laminar decomposition admits such a succinct representation with additive distortion and size depending on the number of hubs. Second,  $k$ -laminar graphs were introduced to model read similarity networks, that are graphs obtained when comparing the short DNA fragments obtained when sequencing several genomes [11]. Such graphs are typically encountered in metagenomic approaches for evolution questions (see e.g. [8]). If graphs appearing in this context often have a laminar structure (a long diameter with low eccentricity), some more complex graphs arise with several laminar structures (see Figure 1 in [11]). Recognizing such complex structure is then related to the problem of binning in metagenomics (see [9]) that consists in sorting DNA fragments into groups that might represent an individual genome or genomes from closely related organisms. Our decomposition algorithm could thus become an interesting tool for the binning problem.

*Related works:* Finding a MESP is NP-complete but can be approximated within a constant factor [4]. Better trade-off between computation time and approximation factor for MESP is obtained in [3]. The problem of representing efficiently the distances in a graph encompasses a vast literature dating from metric embedding [1]. Approximate distance oracles, i.e. compact data-structures for representing an approximation of distances, are investigated in [10]. A particular approach introduced by Peleg [7] resides in assigning a label to each node of a graph such that the distance between two nodes can be estimated from their labels. Several results exist about the trade-off between label size and approximation quality. Exact distance estimation is investigated in [6] and requires  $\Omega(n)$  bits labels for general graphs. Approximation with a constant factor and sub-linear label size is derived in [10]. Some results concern additive approximation such as [5] in the case of hyperbolic graphs.

*Our contributions:* Given a graph that has a hub-laminar decomposition with hub centers sufficiently far apart, we show how to compute an approximated decomposition in polynomial time. The parameters of the computed decomposition (hub radius and covering distance of laminars) are within a constant factor of the optimal decomposition and critical hub centers (those that are extremity of one, three or more laminar paths) are identified up to constant distance error. We also show that such a decomposition allows to construct distance labels with additive distortion and label size proportional to the number of hubs in the decomposition.

## 2 Hub-laminar decomposition

We consider finite, undirected and connected graphs. Given a graph  $G$ , with vertex set  $V(G)$  and edge set  $E(G)$ , we let  $d(u, v)$  denote the *distance* between two vertices, i.e. the length of a shortest path from  $u$  to  $v$ .

Let  $B(u, r) = \{v \in V(G) \mid d(u, v) \leq r\}$  denote the *ball* of radius  $r$  centered at  $u$ . Given a set of vertices  $U$  we denote  $B(U, r) = \cup_{u \in U} B(u, r)$ . Given two sets  $U$  and  $W$  of vertices, we say that  $U$  *k-dominates*  $W$  when every vertex in  $W$  is at distance at most  $k$  from some vertex in  $U$ , i.e.  $W \subseteq B(U, k)$ .

A path  $P$  in  $G$  is a sequence of nodes such that any two consecutive nodes are linked by an edge of  $G$ . We consider only simple paths: a node appears at most once in the sequence. The first node of the sequence and the last one are called the *endpoints* of  $P$ . For the simplicity of notations, we also let  $P$  denote the set of nodes appearing in the sequence. For any vertices  $u$  and  $v$  on  $P$ , we denote by  $P_{uv}$  the subpath of  $P$  having  $u$  and  $v$  as endpoints.

### Definition 1 (Hub-laminar decomposition).

Given a connected undirected graph  $G$ , four positive integers  $r, k, \ell$  and  $\lambda$ , and  $H = \{h_1, \dots, h_q\}$  a set of vertices of  $G$  called hub centers, and  $\mathcal{P} = \{P_1, \dots, P_p\}$  with  $p \leq \lambda$  a set of at most  $\lambda$  paths of  $G$ ; and calling hubs the sets  $(B(h, r))_{h \in H}$  and calling laminars the sets  $(B(P, k))_{P \in \mathcal{P}}$ ;  $(H, \mathcal{P})$  is an  $(r, k, \ell, \lambda)$ -hub-laminar decomposition of  $G$  if the following conditions are satisfied:

1. the laminars link hubs centers: the endpoints of every  $P \in \mathcal{P}$  belong to  $H$
2. the laminars and the hubs cover  $G$ :  $V(G) \subseteq \bigcup_{h \in H} B(h, r) \cup \bigcup_{P \in \mathcal{P}} B(P, k)$
3. hubs centers are  $\ell$  spread:  $\forall h \neq h' \in H, d(h, h') \geq \ell$
4. laminars meet at hubs only: for all  $i \neq j$  and  $uv \in E(G)$  such that  $u \in B(P_i, k)$  and  $v \in B(P_j, k)$ , there is a hub center  $h \in H$  such that  $P_i$  and  $P_j$  both have  $h$  as endpoint and  $u, v \in B(h, r)$
5. the laminars only meet two hubs: Consider a path  $P \in \mathcal{P}$ ,  $h_i, h_j$  the extremities of  $P$  and  $h_z \in H$  such that  $z \neq i, j$  then  $B(P, k) \cap B(h_z, r + 1) = \emptyset$
6. each path  $P \in \mathcal{P}$  is locally a shortest path: considering its endpoints  $h_i$  and  $h_j$ ,  $P_i$  is a shortest path of the graph  $G[B(P, k) \cup B(h_i, r) \cup B(h_j, r)]$

An example is given in Fig. 1. Notice that Axiom 5 says that the graph induced by  $B(P, k) \cup B(h_i, r) \cup B(h_j, r)$ , called the *dumbbell* of  $P$ , is a  $\max(k, r)$ -laminar

graph under the definition of Völkel *et al.* [11]. As a consequence of Axioms 3,  $P$  length must be at least  $\ell$ . Roughly speaking, a hub-laminar decomposition of  $G$  consists in covering  $G$  with at most  $\lambda$  dumbbells of length at least  $\ell$  intersecting near their extremities only. When  $\ell \geq 2r+1$ , a hub-laminar decomposition forms a partition of the edges of  $G$  in the following sense: for each  $uv \in E(G)$ ,  $\exists! h \in H$  s.t.  $u, v \in B(h, r)$  or  $\exists! P \in \mathcal{P}$  s.t.  $u, v \in B(P, k)$ . More generally, any shortest path is included in a laminar or intersects a hub as stated by the following lemma.

**Lemma 1 (shortest path cover).** *Any shortest path  $Q$  either intersects a hub ( $\exists h \in H$  s.t.  $Q \cap B(h, r) \neq \emptyset$ ) or is included in a laminar ( $\exists P \in \mathcal{P}$  with endpoints  $h_1, h_2$  s.t.  $Q \subseteq B(P, k) \setminus B(\{h_1, h_2\}, r)$ ).*

*Proof.* Suppose that no node of  $Q$  is in a hub. According to the cover Axiom 2, all nodes of  $Q$  must be in laminars. Two consecutive nodes  $uv \in Q$  cannot belong to two different laminars since some hub would then contain both  $u$  and  $v$  according to Axiom 4.  $Q$  must then be completely included in the laminar containing its first node.

A hub-laminar decomposition gives naturally raise to a *quotient graph*:

**Definition 2 (quotient graph and reduced quotient).** *Given a graph  $G$  and an  $(r, k, \ell, \lambda)$ -hub-laminar decomposition  $(H, \mathcal{P})$  of  $G$ , the quotient of this decomposition is an edge-labeled multigraph with vertex-set  $H$  and for each  $P \in \mathcal{P}$  joining  $h$  and  $h'$  there is an edge  $hh'$  whose label is the length of  $P$ .*

*The reduced quotient graph of a decomposition  $(H, \mathcal{P})$  is the multigraph obtained from its quotient graph by iteratively removing degree 2 nodes: for every vertex  $u$  of the quotient incident with exactly two edges  $uv$  and  $uw$ , both edges and  $u$  are removed and a new edge  $vw$  is added. It may be a loop if  $v = w$ .*

The number of edges of the quotient is  $|\mathcal{P}| \leq \lambda$  and, since  $G$  is connected, so is the quotient and  $|H| \leq \lambda + 1$ . The size of the quotient and of the reduced quotient are therefore  $O(\lambda)$ . Notice that if the quotient is a cycle, the reduced quotient is empty.

Let the *degree* of a hub be the number of paths its center is the endpoint of. Or equivalently its degree in the quotient graph. We also define the *hub degree* of a node  $h \in H$  as the degree of the corresponding hub.

**Definition 3 (equivalence between decompositions).** *Two hub-laminar decomposition, possibly with different parameters  $r, k, \ell, \lambda$ , are equivalent if they have the same reduced quotient graph, up to an isomorphism  $\phi$  of vertex-sets.*

*Furthermore the decompositions are  $D$ -equivalent if the distance between a hub center and its image is bounded by  $D$ , i.e., for every hub center  $h$   $d(h, \phi(h)) \leq D$ .*

### 3 Distance labeling

Before provide an algorithm for computing a hub-laminar decomposition, we first note that such a decomposition of a graph  $G$  allows to compute a compact representation of distances in  $G$  with additive distortion. A distance labeling is said to be  $c$ -additive and have  $s$  bit labels when the label  $L_u$  assigned to a node  $u$  contains at most  $s$  bits and for all pairs of nodes  $u, v$ , a distance estimation  $\widehat{d}_{uv}$  can be computed from  $L_u$  and  $L_v$  such that  $d(u, v) \leq \widehat{d}_{uv} \leq d(u, v) + c$ .

**Proposition 1.** *Any graph  $G$  with a  $(r, k, \ell, \lambda)$ -hub-laminar decomposition  $(H, \mathcal{P})$  has a  $\max(4k, 2r)$ -additive distance labeling with  $O(\lambda \log n)$  bit labels.*

*Proof.* We assume that hub centers are numbered from 1 to  $q$ ,  $q \leq \lambda$ . For every  $u \in V(G)$ , we define a hub label  $H_u$  consisting in all pairs  $(h, d(u, h))$  for  $h \in H$ . For a node  $u$  in a hub, i.e. when there exists  $h \in H$  such that  $u \in B(h, r)$ , we define its label  $L_u$  as its hub label, i.e.  $L_u := H_u$ . For a node  $u$  in a laminar, i.e. there exists  $P \in \mathcal{P}$  with endpoints  $h_1 < h_2$  such that  $u \in B(P, k) \setminus B(\{h_1, h_2\}, r)$ , we additionally store  $(h_1, h_2, d_P(h_1, u'), d(u', u))$  for some  $u' \in B(u, k) \cap P$  and set  $L_u := (h_1, h_2, d_P(h_1, u'), d(u', u)), H_u$  (we let  $d_P$  denote the distance in the graph induced by  $P$ ).

The distance  $d(u, v)$  between two nodes  $u, v \in V(G)$  is then estimated from their labels  $L_u$  and  $L_v$  as follows. We first compute the estimate through hub centers  $g(u, v) = \min_{h \in H} d(u, h) + d(v, h)$ . If  $L_u$  and  $L_v$  both begin with quadruples  $(h_1, h_2, d(h_1, u'), d(u', u))$  and  $(h'_1, h'_2, d(h'_1, v'), d(v', v))$  respectively with  $h_1 = h'_1$  and  $h_2 = h'_2$ , we detect that  $u$  and  $v$  belong to the same laminar and return the distance estimate  $f(u, v) = \min(g(u, v), g'(u, v))$  where  $g'(u, v) = d(u', u) + |d_P(h_1, u') - d_P(h_1, v')| + d(v', v)$ . Otherwise, we simply return  $f(u, v) = g(u, v)$  as distance estimate.

We now prove that we always have  $d(u, v) \leq f(u, v) \leq d(u, v) + \max(4k, 2r)$ . By triangle inequality, we have  $d(u, v) \leq d(u, h) + d(v, h)$  for all  $h \in H$  and thus obtain  $d(u, v) \leq g(u, v)$ . In the case where  $u$  and  $v$  both belong to the same laminar  $B(P, k)$ , note that  $g'(u, v)$  is the length of a path through vertices  $u', v' \in P$  from  $u$  to  $v$ , implying  $g'(u, v) \leq d(u, v)$ . We thus have  $d(u, v) \leq f(u, v)$  in any case. Now consider a shortest path  $Q$  from  $u$  to  $v$ . First assume  $Q$  intersects a hub: there exists  $h \in H$  such that  $Q \cap B(h, r) \neq \emptyset$ . Consider  $x \in Q \cap B(h, r)$ . We then have  $d(u, v) = d(u, x) + d(x, v) \leq d(u, h) + d(h, x) + d(v, h) + d(h, x) \leq d(u, h) + d(v, h) + 2r$  implying  $g(u, v) \leq d(u, v) + 2r$ . Second, suppose that  $Q$  does not intersect any hub, it must then be included in a laminar according to Lemma 1. Consider  $P \in \mathcal{P}$  with endpoints  $h_1 < h_2$  such that  $Q \subseteq B(P, k) \setminus B(\{h_1, h_2\}, r)$ . Then  $u$  and  $v$  both belong to the laminar and their labels contain quadruples  $(h_1, h_2, d(h_1, u'), d(u', u))$  and  $(h_1, h_2, d(h_1, v'), d(v', v))$  respectively. Consider the sub-graph  $G_P$  induced by  $B(P, k)$ . As  $P$  is a shortest path in  $G_P$ , we have  $d_{G_P}(u', v') \leq d_{G_P}(u, u') + d_{G_P}(u, v) + d_{G_P}(v, v')$ . As  $Q$  is included in  $B(P, k)$  we have  $d(u, v) = d_{G_P}(u, v)$  and we obtain  $|d_P(h_1, u') - d_P(h_1, v')| = d_{G_P}(u', v') \leq d(u, v) + 2k$  and thus get  $f(u, v) \leq g'(u, v) \leq d(u, v) + 4k$ . In any case we have  $f(u, v) \leq d(u, v) + \max(4k, 2r)$ .

## 4 A polynomial time approximation

**Proposition 2.** *Given  $k, r, \ell$  and  $\lambda$ , deciding if a graph  $G$  admits an  $(r, k, \ell, \lambda)$ -hub-laminar decomposition is NP-complete*

*Proof.* The Maximum Eccentricity Shortest Path problem of Dragan and Leitert [4] consists in computing the smallest  $k$  such that  $G$  admits a  $(k, k, 0, 1)$  hub-laminar decomposition. It is an NP-complete problem [4].

Moreover, the best known algorithm [4] for computing a Maximum Eccentricity Shortest Path for a fixed  $k$  takes  $O(mn^{2k+2})$  time, meaning the problem is untractable even for  $\lambda = 1$ . On the other hand, provided parameters  $r, k, \ell$  and  $\lambda$ , it takes  $O(m)$  time to test whether a pair  $(H, \mathcal{P})$  is a  $(r, k, \ell, \lambda)$ -hub-laminar decomposition.

We are therefore now interested in an approximation in the sense that, assuming  $G$  admits a  $(r, k, \ell, \lambda)$ -hub-laminar decomposition, we want to compute a  $D$ -equivalent  $(r', k', \ell', \lambda)$ -hub-laminar decomposition (i.e. with larger parameters but no more hubs nor laminar) in polynomial time, with small  $D$  with respect to the parameters. The  $D$ -equivalence that implies the degree  $\neq 2$  hubs of  $A$  are close to those of  $H$ .

**Theorem 1.** *Consider a graph  $G$  having an  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$  with  $\ell > \max(28r + 17k + 3, 15r + 43k + 3)$  and at least one degree 3 hub. Provided  $K$  and  $R$  with  $K > \max(r + 2k, 2r)$ ,  $R > 4K + 2k$  and  $2R + 5K < \ell - 2r - 9k - 3$ , an  $(R, K, L, \lambda)$ -hub-laminar decomposition  $(A, \mathcal{Q})$ , with  $L \geq \ell - 3R - K + r - 2$ , and which is  $2K$ -equivalent to  $(H, \mathcal{P})$ , can be computed in  $O(\min(n, \lambda)m)$  time.*

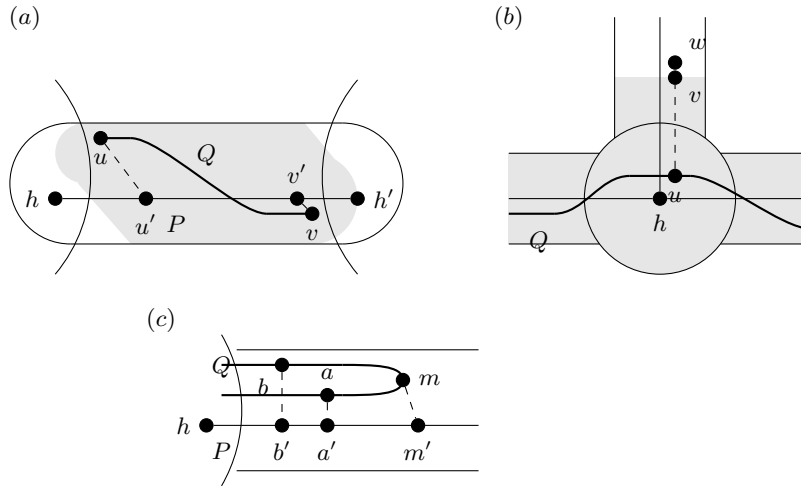
Notice that if  $(H, \mathcal{P})$  has no hub of degree  $\geq 3$ , it is a decomposition into a  $\max(k, r)$ -eccentricity shortest path or into a  $\max(k, r)$ -eccentricity cycle. These two particular cases can be treated separately with similar techniques but are excluded here due to the lack of space. The running time does not rely on the existence of  $(H, \mathcal{P})$ : the algorithm terminates in  $O(nm)$  time on any graph. However, the guarantees on the computed decomposition do rely on it. The proof of this theorem is given in Section 4.7 after giving an algorithm solving the problem and adequate lemmas.

As a corollary, when no value is provided for  $K$  nor  $R$ , it is possible to compute a decomposition with given  $\ell$  (and possibly given  $\lambda$ ) in  $O(n^3m)$  time by trying all possible values of  $R$  and  $K$ .

### 4.1 Algorithm outline

Our algorithm runs in two steps. First *Find\_Hubs* computes the hub centers. It constructs greedily a set of vertices  $A$  initialized by *Find\_Starting\_Hub*. Then *Find\_Laminars* computes the paths between the hub centers using more BFSs.

The key ideas to prove the correctness of all those steps are given in the three corresponding sections. Due to lack of space, the proofs of all Lemmas are detailed in Appendix.



**Fig. 2.** Illustration of the structural properties. (a) Lemma 2: The part of the laminar  $k$ -covered by  $P_{u'v'}$  is  $K$ -covered by  $Q$ . (b) Lemma 3 and Lemma 4. If  $Q$  goes through a hub of degree  $\geq 3$ , a triplet  $u, v, w$  as in Lemma 4 can be found, and it is the only such case far apart the extremities of  $Q$ . (c) Lemma 5. If  $a, m, b$  on  $Q$  and their  $k$ -neighbors on  $P$  are not in the same order, the distance from  $m$  to  $a$  or  $b$  is short.

In the remaining of the paper we consider, sometimes implicitly, a graph  $G$  having a (not known by the algorithm)  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$ , and the decomposition  $(A, \mathcal{Q})$  is the one output by the algorithm detailed below.

## 4.2 Structural properties

The algorithm relies on the selection of hub centers and shortest paths which may not correspond those of  $(H, \mathcal{P})$ . However, three key structural properties ensure that we can rely on them to build a hub-laminar decomposition. The first one shows that any shortest path joining the two hubs of a *dumbbell*  $K$ -covers the central part of the laminar for  $K \geq 3k$ . Finding laminars mainly rely on the following Lemma (see also Fig. 2.a).

**Lemma 2 (Path local covering).** *Consider  $P \in \mathcal{P}$  and a path  $Q$  from  $u$  to  $v$  in the graph induced by  $B(P, k)$ , with  $u', v' \in P$  such that  $d(u, u') \leq k$  and  $d(v, v') \leq k$ . Then every vertex of  $P_{u'v'}$  is at distance at most  $2k$  from  $Q$ .*

*Subsequently, every vertex of  $B(P_{u'v'}, k)$  is at distance at most  $3k$  of  $Q$ .*

The second structural property characterizes vertices of long shortest paths that are near hub centers of the unknown underlying decomposition. The two following lemmas are the key results that allow to find hubs with degree 3 or more (see also Fig. 2.b).

**Lemma 3 (Hub in the middle).** *Consider  $K \geq 3k$  and let  $Q$  be a shortest path from  $a$  to  $b$  and  $u \in V(Q)$  at distance greater than  $K + 6k$  from  $a$  and  $b$ .*



If there exists  $vw \in E(G)$  such that  $d(u, v) = K$  and  $d(Q, w) = K + 1$ , then there exists a hub center  $h \in H$  with  $d(u, h) \leq K + r$ .

**Lemma 4 (Degree  $\geq 3$  Hub Detection).** Consider  $K \geq 3k$  and let  $Q$  be a shortest path from  $a$  to  $b$ , with  $u \in V(Q)$  at distance more than  $r + 4K + 9k + 2$  from  $a$  and  $b$ .

If there exists a hub  $h \in H$  of degree at least 3 with  $d(u, h) \leq K$ , then there exists  $vw \in E(G)$  and a vertex  $x \in Q$  at distance more than  $K + 6k$  from  $a$  and  $b$  such that  $d(x, v) = K$ ,  $d(Q, w) = K + 1$ .

Finally, the third structural property implies that a shortest path entering and leaving a lamina by the same hub may not go deep inside as implied by the following lemma (see also Fig. 2.c).

**Lemma 5 (Bounded zig-zag).** Consider a shortest path  $Q$  in the graph induced by  $B_G(P, k)$  with  $P \in \mathcal{P}$  and three successive nodes  $a, m, b$  on  $Q$  with  $a', m', b'$  on  $P$  such that  $d_G(a, a') \leq k$ ,  $d_G(m, m') \leq k$ ,  $d_G(b, b') \leq k$ .

If  $a'$  is between  $b'$  and  $m'$  ( $a' \in P_{m', b'}$ ), then we have  $d_G(a, m) \leq 3k$ .

### 4.3 The stopBFS function

Let  $G$  be a vertex-colored graph with some uncolored vertices. The **stop-BFS** procedure, provided a vertex  $d$  and a color  $c$ , consists in running a usual Breadth-First Search, starting at vertex  $d$ , and returning a node  $f$  and a path  $P$  from  $d$  to  $f$ , with the following additional rules:

- only vertices without color  $c$  put in the BFS queue,
- the BFS stops immediately if a vertex  $f$  is visited (i.e. extracted from BFS queue) such that  $f$  has a colored neighbor whose color is not  $c$ ,
- otherwise, it stops because the queue is empty, let  $f$  be the last visited vertex,
- function  $stopBFS(d, c)$  returns  $f$  and the BFS path  $P$  from  $d$  to  $f$  ( $P$  is a shortest path in the graph induced by  $G$  after removing  $c$ -colored vertices).

### 4.4 Computing the hub centers

Function *Find\_Hubs* works by coloring vertices and adding some of them to a set  $A$  of hubs. Initially all vertices are uncolored. A vertex inside  $B(a, R)$  for  $a \in A$  gets a new color  $col(a)$ . Vertices supposed to belong to a lamina but to no hub are colored with color  $lam$ . Given a path  $P$ ,  $r3K(P)$  denotes the subpath of  $P$  obtained by removing the  $3K$  first and  $3K$  last vertices of  $P$ . Each time a vertex  $a$  is inserted in  $A$ , the vertices of  $B(a, R)$  are colored with a color identifying the hub. While there are uncolored nodes in  $G$ , a stop-BFS (see Section 4.3) rooted near  $a \in A$  is run. In the pseudo-code given on next page, five cases may occur:

- (a). Either we stop very near the starting hub (Line 5), closer than the parameter  $L$ , meaning that we are in a disconnected part of an already visited lamina.
- (b). Or (Line 8) there is Lemma 3 configuration (*hub in the middle*) implying that there is a hub with degree  $\geq 2$  close to the node  $h$  added to  $A$ .

- (c). Or the BFS comes back to the starting hub without finding such an enlargement (Line 12). A degree 2 hub is then added in the middle of the path found that is considered to correspond to two parallel laminars.
- (d). Or we stop in a dead end (Line 17) meaning that there is a degree 1 hub near  $f$  which is added to  $A$ .
- (e). Or we stop near another already discovered hub (Line 20) meaning that there is no new nearby hub and we just traversed a lamina that is now colored.

```

1 Find_Hubs
   Input: A graph  $G$ , integers  $R$  and  $K$ , a starting node  $s$ .
   Output: A set of vertices  $A$ .
2 Set  $A = \{s\}$ ; color every vertex in  $B(s, R)$  with color  $col(s)$ 
3 While  $\exists a \in A$  and an uncolored vertex  $d \in G$  such that  $d(a, d) = R + 1$  do
4   Let  $f, P = stopBFS(d, col(a))$  ( $f$  is the last vertex of  $P$ ).
5   If  $P$  length is at most  $2R + 4K + 2$  /* Case (a) */
6   then
7     Color all vertices visited by  $stopBFS(d, col(a))$  with color  $lam$ .
8   else if  $\exists w, h$  s.t.  $col(w) \neq col(a)$  and  $h \in r3K(P)$  and  $d(w, h) = K + 1$  and
9      $d(w, P) = K + 1$  /* Case (b) */
10    then
11      Let  $h$  be the first vertex of  $r3K(P)$  satisfying the above condition.
12      Add  $h$  to  $A$  and color every vertex in  $B(h, R)$  with color  $col(h)$ .
13    else if  $f$  is at distance at most  $2K$  from  $B(a, R)$  /* Case (c) */
14    then
15      Let  $m$  be a vertex in the middle of  $P$ .
16      Add  $m$  to  $A$  and color every vertex in  $B(m, R)$  with color  $col(m)$ .
17      Color uncolored vertices in  $B_{G \setminus \{B(d, R) \cup B(f, R)\}}(P, K)$  with color  $lam$ .
18    else if  $f$  is not adjacent to a colored vertex /* Case (d) */
19    then
20      Add  $f$  to  $A$  and color every vertex in  $B(f, R)$  with color  $col(f)$ .
21    else /* Case (e) */
22      Color uncolored vertices in  $B_{G \setminus \{B(d, R) \cup B(f, R)\}}(P, K)$  with color  $lam$ .

```

The *Find\_Hubs* procedure finds the hubs of  $(H, \mathcal{P})$  up to those of degree 2. More precisely, Lemmas 6 and 9 imply that there exists a bijection between Degree 1 hubs of  $(H, \mathcal{P})$ , and vertices of  $A$  selected at line 19, and another one between hubs of  $(H, \mathcal{P})$  with degree at least 3, and vertices of  $A$  selected at line 11. Moreover, the distance between the vertex of  $A$  and the corresponding vertex of  $H$  is at most  $2K$ . The order of the lemmas is determined by the way they depend on each other in the proofs.

**Lemma 6 (Computed hubs of degree  $\neq 2$  are close to those of  $H$ ).**  
*Consider a graph  $G$  having a  $(r, k, \ell, \lambda)$  hub-lamina decomposition  $(H, \mathcal{P})$  with*

$2R + 5K < \ell - 2r - 9k - 3$ , and a hub center  $h_0 \in H$  with hub degree three or more. Suppose that *Find\_Hubs* is called with a starting node  $s$  such that  $s$  is at distance at most  $K + r$  from a hub center  $h$  with hub degree at least 3, then:

(i) For every vertex  $a$  added in  $A$  at line 11, there is a hub center  $h \in H$  of hub degree at least 2 at distance at most  $K + r$  from  $a$ .

(ii) For every vertex  $a$  added in  $A$  at line 19, there is a hub center  $h \in H$  of hub degree 1 at distance at most  $2(k + r)$  from  $a$ .

The following lemma states that vertices of  $A$  are far apart.

**Lemma 7 (Computed hubs are far apart).** Consider a graph  $G$  having a  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$  with  $2R + 5K < \ell - 2r - 9k - 3$  such that at least one vertex  $h \in H$  has hub degree three or more.

Suppose that *Find\_Hubs* is called with a starting node  $s$  such that  $s$  is at distance at most  $R - r$  from a hub center  $h \in H$  of hub degree at least 3.

The algorithm *Find\_Hubs* outputs a set  $A$  of vertices that are at distance at least  $\ell - 3R - K + r - 2$  one from another.

The following lemma imply the termination of *Find\_Hubs*.

**Lemma 8 (Uncolored vertices are close to  $A$ ).** Consider a graph  $G$  having a  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$  with  $2R + 5K < \ell - 2r - 9k - 3$  such that at least one hub has degree three or more.

Suppose that *Find\_Hubs* is called with a starting node  $s$  such that  $s$  is at distance at most  $R - r$  from a hub center  $h \in H$  of hub degree at least 3.

The algorithm *Find\_Hubs* ends with every vertex colored or at distance  $3R + 2K + 1$  at most from a vertex  $a$  of  $A$ .

**Lemma 9 (Hubs of  $H$  of degree  $\neq 2$  are close to computed ones).** Consider a graph  $G$  having a  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$  with  $2R + 5K < \ell - 2r - 9k - 3$  such that at least one hub of degree three or more.

Suppose that *Find\_Hubs* is called with a starting node  $s$  such that  $s$  is at distance at most  $K + r$  from an hub center  $h \in H$  of hub degree at least 3, then after termination of *Find\_Hubs*:

(i) For every hub  $h \in H$  of degree at least 3, there exists a vertex  $a$  in  $A$  at distance at most  $K + r$  from  $h$ .

(ii) For every hub  $h \in H$  of degree 1, there exists a vertex  $a$  in  $A$  at distance at most  $2(k + r)$  from  $h$ .

#### 4.5 Computing the first hub center

To compute a starting node  $s$  at distance at most  $K + r$  from a degree 3 hub, we use a procedure similar to *Find\_Hubs* but starting from an arbitrary start vertex  $r$  until a hub is detected according to Line 10 (case (c)). The only case where the procedure may fail is when there is only one such hub and  $r$  was chosen sufficiently close to it. In that case, running again the procedure from the last visited node from  $r$  allows to find it.

## 4.6 The function to detect laminars

Since *Find\_Hubs* constructs the set of hub centers, at this step we have the hubs. We just have to identify the laminars and their paths. Each path is found by a BFS starting at an hub center and ending at the first hub center encountered. Then we remove from the graph the vertices from the laminar, but not the hubs. The process ends when the graph consists in disconnected hubs only.

### 1 Find\_Laminars

**Input:** A graph  $G$ , integers  $R$  and  $K$ , a set  $A$  of hub centers.

**Output:** a hub-laminar decomposition  $(A, \mathcal{Q})$ .

2  $\mathcal{Q} = \emptyset$

3 Mark all vertices as deletable

4 **For each** vertex  $a$  in  $A$  **do**

5     | Mark the vertices in  $B(a, R)$  as undeletable

6 **While** there exists  $a \in A$  such that  $B(a, R + 1) \neq B(a, R)$  **do**

7     | Run a BFS starting at  $a$  and stopping on the first vertex  $a' \in A$ ,  $a' \neq a$

8     | Add to  $\mathcal{Q}$  the path  $Q$  from  $a$  to  $a'$  computed by this BFS

9     | Delete from  $G$  the deletable vertices from  $B(Q, K)$

**Lemma 10.** *Consider a graph  $G$  having a  $(r, k, \ell, \lambda)$  hub-laminar decomposition  $(H, \mathcal{P})$  with  $2R + 5K < \ell - 2r - 9k - 3$  with at least one hub of degree three or more. Suppose that *Find\_Laminars* is called with a set  $A$  such that for every hub center  $h \in H$  of hub degree different from two, there is a vertex  $a$  in  $A$  at distance  $2K$  at most from  $h$ . If there exists an edge  $(u, v) \in E(G)$  with  $u \in B(Q, K)$  and  $v \notin B(Q, K)$  for some path  $Q$  with endpoints  $a$  and  $a'$  that is return by the algorithm, then we have  $v \in B(\{a, a'\}, R)$ .*

Lemma 10 implies that *Find\_Laminars* terminates. Indeed, if  $G$  still contains deletable vertices when the algorithm stops. The connexity of  $G$  implies that there exist a pair  $(u, v)$  of vertices such that  $v$  is an undeleted deletable vertex and  $u$  is not. The lemma then ensures that  $v$  should indeed be undeletable, raising a contradiction.

## 4.7 Proof of Theorem 1

Let us first show that the output of the algorithm fulfills the definition of a hub-laminar decomposition.

- Axiom 1. The laminars link hubs centers: The set of paths  $\mathcal{Q}$  is defined in the function *Find\_Laminars*. They are by definition paths between two vertices of  $A$ , the set of hubs returned by the previous function.
- Axiom 2 (the laminars and the hubs cover  $G$ ). This is a result of Lemma 10 as discussed before.

- Axiom 3. Hubs centers are  $L$  far apart: This is a direct result of lemma 7.
- Axiom 4 (laminars meet at hubs only). This is a direct consequence of Lemma 10.
- Axiom 5 (the laminars only meet two hubs). Let  $Q$  be a path of  $\mathcal{Q}$  with end points  $a_1, a_2 \in A$ . Assume the existence of a third vertex  $a_3 \in A$  such that  $B(a_3, R)$  intersects  $B(Q, K)$ , i.e.  $a_3$  is at distance  $K + R$  at most from some vertex  $q \in Q$ . We then have:

$$\begin{aligned} d(a_1, a_3) &\leq d(a_1, q) + K + r \\ d(q, a_2) &\geq d(a_2, a_3) - d(q, a_3) \geq L - K - r \\ d(a_1, a_2) &= d(a_1, q) + d(q, a_2) \geq d(a_1, q) + L - K - r > d(a_1, a_3) \end{aligned}$$

Without loss of generality, assume that  $Q$  started on  $a_1$ ,  $Q$  is then the shortest path between  $a_1$  and any vertex of  $A$  in the remaining graph when computing  $Q$ . As  $a_3$  is not visited before  $a_2$ , some vertices between  $q$  and  $a_3$  must have been deleted during a previous step of the algorithm. This means that  $Q$  meets an other laminar outside a hub, in contradiction with Axiom 4 that we have already established.

- Axiom 6. Each path  $Q \in \mathcal{Q}$  is locally a shortest path: each path  $Q$  with endpoint  $a, a'$  is a shortest path of the remaining graph when computing  $Q$  which contains the dumbbell  $B(Q, K) \cup B(a, R) \cup B(a', R)$ .

The  $2K$ -equivalence is a consequence of Lemma 6 and Lemma 9, which allow to build the bijection  $\phi$  between hub centers with hub degree different from 2. Notice  $2K \geq \max(K + r, 2(k + r))$ . Decomposition  $(A, \mathcal{Q})$  has  $\lambda$  hubs at most since it has no more degree 2 hubs than  $(H, \mathcal{P})$ . Our algorithm indeed adds degree 2 hubs in two cases only. First, when the conditions of Lemma 3 (*hub in the middle*) are met at Line 8, the added hub is associated by Lemma 6 to a hub of  $H$ . Second, when we encounter a self loop in the reduced quotient, i.e. a sequence of (at least 2) laminars of  $(H, \mathcal{P})$  connected by (at least 1) hubs of degree 2, the algorithm then adds only one hub (at Line 14 according to case (c)).

Regarding the time complexity, apart from case (a), each iteration of the *while* loop in *Find\_Hubs* corresponds to finding a hub or a laminar. There are thus  $O(|A| + |\mathcal{Q}|)$  such iterations, and their overall cost is  $O(\min(\lambda, n)m)$ . In the iterations corresponding to Case (a), all vertices visited by StopBFS are colored: the overall cost of such iterations is thus  $O(m)$ . Similarly, *Find\_Laminars* consists in  $\lambda$  iterations costing  $O(m)$  each.

## 5 Acknowledgments

The authors thank Michel Habib for inspiring discussions about  $k$ -laminar graphs, and Eric Bapteste, Philippe Lopez and Chloé Vigliotti for raising the problem of identifying complex laminar structures in biological graphs.

## References

1. Assouad, P.: étude d'une dimension métrique liée à la possibilité de plongements dans  $\mathbf{R}^n$ . C. R. Acad. Sci. Paris Sér. A-B 288(15), A731–A734 (1979)
2. Badoiu, M., Chuzhoy, J., Indyk, P., Sidiropoulos, A.: Low-distortion embeddings of general metrics into the line. In: Gabow, H.N., Fagin, R. (eds.) Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005. pp. 225–233. ACM (2005), <http://doi.acm.org/10.1145/1060590.1060624>
3. Birmelé, E., de Montgolfier, F., Planche, L.: Minimum eccentricity shortest path problem: An approximation algorithm and relation with the k-laminarity problem. In: Combinatorial Optimization and Applications - 10th International Conference, COCOA 2016, Hong Kong, China, December 16-18, 2016, Proceedings. pp. 216–229 (2016), [http://dx.doi.org/10.1007/978-3-319-48749-6\\_16](http://dx.doi.org/10.1007/978-3-319-48749-6_16)
4. Dragan, F.F., Leitert, A.: On the minimum eccentricity shortest path problem. In: Algorithms and Data Structures - 14th International Symposium, WADS 2015, Victoria, BC, Canada, August 5-7, 2015. Proceedings. pp. 276–288 (2015), [http://dx.doi.org/10.1007/978-3-319-21840-3\\_23](http://dx.doi.org/10.1007/978-3-319-21840-3_23)
5. Gavoille, C., Ly, O.: Distance labeling in hyperbolic graphs. In: Deng, X., Du, D. (eds.) Algorithms and Computation, 16th International Symposium, ISAAC 2005, Sanya, Hainan, China, December 19-21, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3827, pp. 1071–1079. Springer (2005), [http://dx.doi.org/10.1007/11602613\\_106](http://dx.doi.org/10.1007/11602613_106)
6. Gavoille, C., Peleg, D., Pérennes, S., Raz, R.: Distance labeling in graphs. J. Algorithms 53(1), 85–112 (2004), <http://dx.doi.org/10.1016/j.jalgor.2004.05.002>
7. Peleg, D.: Proximity-preserving labeling schemes. Journal of Graph Theory 33(3), 167–176 (2000), [http://dx.doi.org/10.1002/\(SICI\)1097-0118\(200003\)33:3<167::AID-JGT7>3.0.CO;2-5](http://dx.doi.org/10.1002/(SICI)1097-0118(200003)33:3<167::AID-JGT7>3.0.CO;2-5)
8. Saw, J.H., Spang, A., Zaremba-Niedzwiedzka, K., Juzokaite, L., Dodsworth, J.A., Murugapiran, S.K., Colman, D.R., Takacs-Vesbach, C., Hedlund, B.P., Guy, L., et al.: Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. Phil. Trans. R. Soc. B 370(1678), 20140328 (2015)
9. Thomas, T., Gilbert, J., Meyer, F.: Metagenomics-a guide from sampling to data analysis. Microbial informatics and experimentation 2(1), 3 (2012)
10. Thorup, M., Zwick, U.: Approximate distance oracles. J. ACM 52(1), 1–24 (2005), <http://doi.acm.org/10.1145/1044731.1044732>
11. Völkel, F., Baptiste, E., Habib, M., Lopez, P., Vigliotti, C.: Read networks and k-laminar graphs (Mar 2016), <https://hal.inria.fr/hal-01282715>, arXiv:1603.01179

## Appendix : Proofs

### Proof of Lemma 2

Let us define  $x_0 = u$ ,  $x_s = v$  and  $Q = x_0, \dots, x_s$ .

The second assertion of the lemma is straightforward given the first one. To prove the latter, we define, for all  $l$  between 0 and  $s$ , the subpath  $Q_l = x_0, x_1, \dots, x_l$  and  $x'_l$  in  $P$  such that  $d(x_l, x'_l) \leq k$

Let us show by induction on  $l$  that every vertex of  $P$  between  $u'$  and  $x'_l$ , is at distance at most  $2k$  of  $P_l$ .

- For  $l = 0$ ,  $Q_0 = x_0 = u$  and  $x'_0 = u'$ . As  $u'$  is at distance  $k$  of  $u$ , the result is true for  $l = 0$ .

- Let  $l$  in  $(1 \dots s)$  such that the property is verified for  $l - 1$ .

Every vertex  $y$  of  $P_{u'x'_{l-1}}$  is at distance at most  $2k$  of  $P_{l-1}$  by induction hypothesis, and thus at distance at most  $2k$  of  $P_l$ .

Moreover, by the triangle inequality:

$$d(x'_{l-1}, x'_l) \leq d(x'_{l-1}, x_{l-1}) + d(x_{l-1}, x_l) + d(x_l, x'_l) \leq 2k + 1 \quad (1)$$

As the sub-path of  $P$  between  $x'_{l-1}$  and  $x'_l$  is a shortest path, it follows that, for every vertex  $y$  of  $P_{x'_{l-1}x'_l}$ ,

$$d(x'_{l-1}, y) \leq k \text{ or } d(x'_l, y) \leq k, \quad (2)$$

meaning that  $y$  is at distance at most  $2k$  of  $P_{l-1}$  or of  $x_l$ .

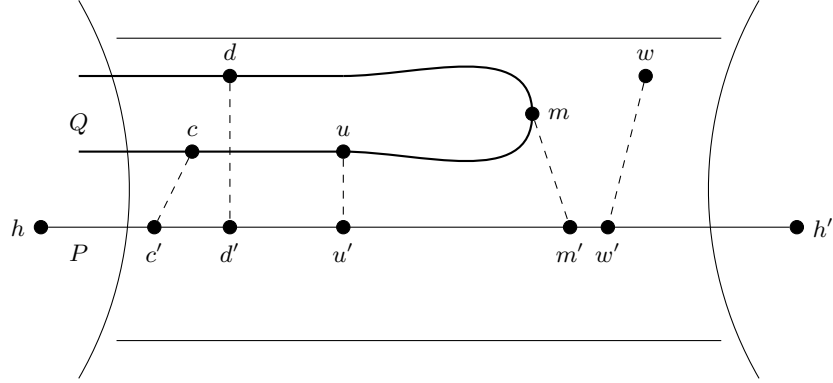
The property is verified by induction, and the lemma follows for  $l = s$ .

### Proof of Lemma 3

For the sake of contradiction, suppose no such hub exists. Then  $u$  must be in a laminar with path  $P$  from  $h$  to  $h'$  and there exists  $u' \in P$  such that  $d(u, u') \leq k$ .

Suppose first that  $w \in B(P, k)$ . Consider  $w' \in P$  such that  $d(w, w') \leq k$ , and suppose w.l.o.g. that  $u' \in P_{hw'}$ . Suppose there exist  $z \in Q$  and  $z' \in P_{w'h'}$  with  $d(z, z') \leq k$ . Lemma 2 applied to  $Q_{uz}$  then implies that  $d(w, Q) \leq 3k \leq K$ , which is a contradiction. Thus,  $Q$  cannot intersect  $B(P_{w'h'}, k)$ .

We can thus define  $m'$  as the  $k$ -neighbor of  $Q$  on  $P_{u'w'}$  furthest from  $u'$  and  $m \in Q$  such that  $d(m, m') \leq k$ , as shown in Figure 3. As  $d(w, Q) > 3k$ ,  $d(w', m') \geq k + 1$  so that, using  $d(u', w') \leq K + 2k + 1$ ,  $d(u', m') \leq K + k$ . It implies that  $d(u, m) \leq K + 3k$ . The hypothesis on  $d(u, a)$  and  $d(u, b)$  then imply that there exist two vertices  $c$  and  $d$  on  $Q$  at distance  $3k + 1$  of  $m$ . As  $u$  is at distance greater than  $K + r$  of any hub-center,  $Q_{c,d} \subset B(P, k)$ . Thus, the fact that  $m'$  is the furthest  $k$ -neighbor of  $Q$  on  $P$  contradicts Lemma 5 applied to  $c$ ,  $m$  and  $d$ .



**Fig. 3.** Illustration of the proof of Lemma 3

We therefore have  $w \notin B(P, k)$ . Consider then the last node  $x$  in  $B(P, k)$  on the path from  $u$  to  $v$ . By assertion 4. of the definition of a hub-laminar decomposition, there exist a hub center  $h$  such that  $x \in B(h, r)$ , and thus  $d(u, h) \leq K + r$ .

#### Proof of Lemma 4

Consider three paths  $P_i, P_k, P_l$  of  $\mathcal{P}$  with  $h$  as an endpoint and vertices  $x'_i, x'_j, x'_l$  on those paths, each at distance  $r + K + 3k + 2$  from  $h$ .

Assume first that those three vertices are at distance at most  $K$  of respectively  $x_i, x_j, x_l$ , vertices of  $Q$ . None of the last three vertices belongs to the hub  $B(h, r)$  as  $d(h, x_i) \geq d(h, x'_i) - d(x'_i, x_i) \geq r + 3k + 2$ . Moreover, we may assume w.l.o.g that  $x_j, x_i$  and  $x_l$  are in that order in  $Q$ . There exist therefore a maximal subpath  $Q_{ce}$  of  $Q$  that is part of  $B(P_i, k) \setminus B(h, r)$  and that contains  $x_i$ .

Let  $c'$  and  $d'$  be vertices of  $P_i$  such that  $d(c, c') \leq k$  and  $d(d, d') \leq k$ . Then  $d(h, c') \leq d(h, c) + k \leq r + k + 1$  and similarly for  $d'$ . As  $d(h, x'_i) > r + k + 1$ , Lemma 5 applies to  $c, x_i$  and  $e$  and implies that  $d(c, x_i) \leq 3k$  or  $d(d, x_i) \leq 3k$ . In both cases, as  $d(h, c) = d(h, d) = r + 1$  and  $d(x_i, x'_i) \leq K$ ,  $d(h, x'_i) \leq r + K + 3k + 1$ , which is a contradiction.

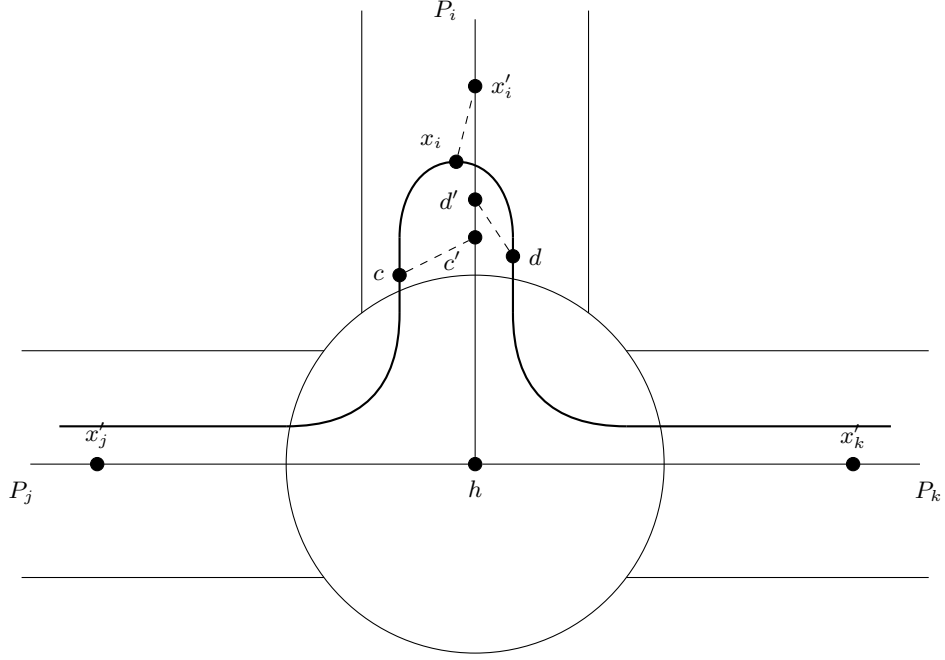
One of the three vertices  $x'_i, x'_j$  or  $x'_k$  is therefore at distance more than  $K$  from  $Q$ , for instance  $x'_i$ . When following  $P_i$  from  $h$  to  $x_i$ , let  $v$  be the last vertex at distance  $K$  from  $Q$ ,  $w$  the following vertex of  $P_i$  and  $x$  a vertex of  $Q$  such that  $d(x, v) = K$ . Then  $d(Q, w) = K + 1$  and, assuming w.l.o.g that  $x \in Q_{u,b}$ ,

$$d(h, v) \leq r + K + 3k + 2 \quad (3)$$

$$d(u, x) \leq d(u, h) + d(h, v) + d(v, x) \leq r + 3K + 3k + 2 \quad (4)$$

$$d(x, b) = d(u, b) - d(u, x) \geq K + 6k \quad (5)$$





**Fig. 4.** Illustration of the proof of Lemma 4

**Proof of Lemma 5**

The following relations derive easily from the hypothesis:

$$d_G(b', m') = d_G(b', a') + d_G(a', m') \quad (6)$$

$$d_G(b, a) = d_G(b, m) + d_G(m, a) \quad (7)$$

$$d_G(b, a) \leq d_G(b', a') + 2k \quad (8)$$

$$d_G(b', m') \leq d_G(b, m) + 2k \quad (9)$$

$$d_G(m, a) \leq d(m', a') + 2k \quad (10)$$

It follows from equations 7 and 8 that

$$d_G(m, a) \leq d_G(b', a') + 2k - d_G(b, m) \quad (11)$$

Using Equation 6,

$$d_G(m, a) \leq d_G(b', m') - d_G(a', m') + 2k - d_G(b, m) \quad (12)$$

Using Equation 9,

$$d_G(m, a) \leq 4k - d_G(a', m') \quad (13)$$

Using Equation 10,

$$d_G(m, a) \leq 6k - d_G(m, a) \quad (14)$$

Finally, we get,

$$d_G(m, a) \leq 3k \quad (15)$$

**Proof of Lemma 6**

(i) This is a direct result of lemma 3.

(ii) Let  $f$  be added to  $A$  at line 19.

By induction,  $d$  is in a lamina  $L$  with endpoints  $h$  and  $h'$ , such that  $B(h, r)$  is in  $B(a, R)$ . Therefore  $P_{d,f}$  is such that  $f$  is in  $L$  or  $Q$  goes through  $h'$ .

Assume  $h'$  of degree at least 2. Then there exists a second lamina  $L'$  incident to  $h'$ , with path  $P' = (h' = v'_0, v'_1, \dots, v'_l)$ . Moreover,

$$d(d, f) \geq d(d, v'_l) \geq d(d, h') + d(h', v'_l) - 2r \geq d(d, h') + l - 2r \quad (16)$$

so that

$$d(h', f) \geq d(d, f) - d(d, h') \geq l - 2r \quad (17)$$

As any path from  $d$  to  $f$  has a vertex in  $B(h', r)$ ,  $h'$  is not of degree at least 3, as otherwise Lemma 4 ( $K \geq 2r$ ) would imply that  $h'$  should have been detected at line 11.

By induction we deduce that there is no hub of degree more than 3 in  $G \setminus B(a, R)$  and that any path in  $G \setminus B(a, R)$  starting on  $d$  goes through the same set of hubs in the same order. This set either end with an hub of degree 1 or with the hub  $h$  (the last lamina intersects  $B(a, R)$ ).

In the second case, let  $v_0, v_1, \dots, v_z$  be the path associated with the lamina, and  $v_j$  the vertex closest to  $v_z$  and not in  $B(a, R)$ .

$$d(d, v_j) \geq d(d, v_0) + d(v_0, v_j) - 2r \quad (18)$$

$$d(d, f) \leq d(d, v_f) + k \leq d(d, v_0) + d(v_0, v_f) + k \quad (19)$$

$$d(d, v_0) + d(v_0, v_f) + k \geq d(d, v_0) + d(v_0, v_j) - 2r \quad (20)$$

$$d(v_0, v_f) \geq d(v_0, v_j) - 2r - k \quad (21)$$

$$d(v_f, a) \leq R + 2r + k \quad (22)$$

$$d(f, a) \leq R + 2r + 2k \quad (23)$$

In the first case, let  $h''$  be the ending hub of degree 1 and  $v''_0, v''_1, \dots, h''$  be the path associated with the lamina.

If  $v''_0 = h$  ( $P_{d,f}$  doesn't intersect an hub of degree 2) then :

$$d(d, f) \leq d(v_a, v_f) + 2k \quad (24)$$

$$d(d, f) \geq d(d, h'') \geq d(v_a, h'') - k \quad (25)$$

$$d(v_a, v_f) + 3k \geq d(v_a, h'') \quad (26)$$

$$d(f, h'') \leq 4k \quad (27)$$

Else,

$$d(d, f) \leq d(d, v_0'') + d(v_0'', v_f) + k \quad (28)$$

$$d(d, f) \geq d(d, h'') \geq d(d, v_0'') + d(v_0'', h) - 2r \quad (29)$$

$$d(v_0'', v_f) + k + r \geq d(v_0'', h) \quad (30)$$

$$d(f, h'') \leq 2k + 2r \quad (31)$$

### Proof of Lemma 7

Let  $a, a'$  be vertices added to  $A$  at line 19 or line 11, by lemma 6 they are at distance at most  $2K$  of some hubs  $h, h'$ . If  $h \neq h'$ , then  $h$  and  $h'$  are distant of at least  $l$  and  $a, a'$  distant of at least  $l - 4K$ , we have the result.

Assume  $h = h'$  and  $a$  was added first in  $A$ . Every vertex at distance  $r + K$  of  $h$  was colored with color  $col(a)$ , in particular  $a'$ . Furthermore,  $a'$  is a vertex computed with the function *stopBFS*, this contradicts the fact that *stopBFS* returns a path of vertices uncolored or with color *lam*.

To prove the rest of the lemma, let us show the following :

*Claim:* For every vertex  $m$  added in  $A$  at line 14,  $m$  is distant of at least  $l - 3R - K + r - 2$  of any other vertex of  $A$ .

Let  $m$  be a vertex added in  $A$  at line 14,  $P_{d,f}$  such that  $m$  was selected as the middle of the path.

**Case 1:  $P_{d,f}$  doesn't meet an hub.** Then  $d$  and  $f$  are in the same lamina  $L$  with path  $P = v_0, v_1, \dots, v_z$ . Let  $a$  be the vertex of  $A$  such that  $d$  is at distance  $R + 1$  of  $a$ , we have

$$d(d, f) \leq d(d, a) + d(a, f) \leq 2R + 2 + 2K \quad (32)$$

Let  $v_f$  (resp.  $v_d$ ) be a vertex of  $P$  at distance  $k$  of  $f$  (resp.  $d$ ). Assume that the shortest path of length  $2R + 2 + 2K$  is in  $L$ , then

$$d(v_d, v_f) \leq d(v_d, d) + d(d, f) + d(f, v_f) \leq 2R + 2 + 2K + 2k \quad (33)$$

Else, the path goes through a hub, meaning that  $d$  and  $f$  are at distance less than  $2R + 2 + 2K + r$  of an hub center. W.l.o.g, assume that the hub center is  $v_0$ .

$$d(v_0, v_d) \leq d(v_0, d) + d(d, v_d) \leq 2R + 2 + 2K + r + k \quad (34)$$

$$\text{Similarly, } d(v_0, v_f) \leq 2R + 2 + 2K + r + k \quad (35)$$

$$\text{So that } d(v_d, v_f) \leq 2R + 2 + 2K + r + k \quad (36)$$

In both cases  $v_d$  and  $v_f$  are at distance at most  $2R + 2 + 2K + r + k$ .

Let  $v_m$  be a vertex of  $P$  at distance at most  $k$  of  $m$ . If the vertices  $v_p, v_m$  and  $v_d$  are not in that order in  $P$ ,  $m$  is at distance at most  $3k$  of  $f$  and  $d$  by lemma 5. This contradicts the fact that  $P_{d,f}$  is of length at least  $2R + 4K + 2$ .

If  $v_f$ ,  $v_m$  and  $v_d$  are in that order in  $P$ ,

$$d_{B(P_{v_d, v_m, k})}(d, m) \leq d(d, v_d) + \frac{d(v_d, v_f)}{2} + d(v_m, m) \leq \frac{2R + 2 + 2K + r + k}{2} + 2k \quad (37)$$

$$\text{Similarly, } d_{B(P_{v_f, v_m, k})}(f, m) \leq \frac{2R + 2 + 2K + r + k}{2} + 2k \quad (38)$$

As  $B(P_{v_d, v_m, k})$  is included in  $B(P_{d, m}, K)$ , Lemma 2 implies that the length of  $P_{d, m}$  is at most  $R + 1 + K + (r + k)/2 + 2k$ , leading again to a contradiction.

**Case 2:  $P_{d, f}$  meets an hub  $h'$ .** Assume that  $h'$  is at distance  $r$  of  $P_{f, m}$ . Then

$$d(d, m) \geq d(f, h') + d(h', m) - 2r \quad (39)$$

and thus

$$\begin{aligned} d(a, m) &\geq d(f, m) - d(a, h) - d(h, f) \\ &\geq d(f, h') + d(h', m) - 2r - d(a, h) - d(h, f) \\ &\geq d(h, h') - d(h, a) - d(a, f) + d(h', m) - 2r - d(a, h) - d(h, f) \\ &\geq l - 3R - K + r - 2 \end{aligned}$$

A symmetric reasoning can be done if  $h'$  is at distance  $r$  of  $P_{d, m}$ .

As any path from  $m$  to a vertex of  $G \setminus B(P_{d, f}, K)$  goes through  $B(a, R + 4K)$ , we have the final result.

### Proof of Lemma 8

Let  $y$  be an uncolored vertex at distance more than  $3R + 4K + 1$  of  $a$  and  $x_0$  the closest colored vertex. We denote by  $Q = (x_0, x_1, \dots, x_t = y)$  the shortest path from  $x_0$  to  $y$ . All vertices in  $Q$  but  $x_0$  are then uncolored.

$x_0$  cannot have been colored by a vertex of  $A$ , as  $x_1$  would have been selected at line 3 and thus colored.  $x_0$  has therefore color  $lam$ . It cannot have been colored at Line 7 as  $y$  would then also be colored. It has therefore been colored by a path  $P_{d, f}$ .

$x_0$  is at distance  $K$  of some vertex  $u$  in  $P_{d, f}$ , otherwise  $x_1$  would have been colored by  $u$ . It implies that  $u$  is at distance less than  $3K$  of  $d$  or  $f$ , otherwise  $x_1$  would have been colored at line 11. We may suppose w.l.o.g that it is at distance at most  $3K$  from  $d$ , so that  $d(d, x_0) \leq 4K$ .

We have that  $d$  is at distance  $R + 1$  of some  $a \in A$ . Vertices added to  $A$  at line 15 have no uncolored vertex at distance  $R + 1$ . Indeed, Lemmas 3 and 7 imply that a vertex would then have been added at line 11. By Lemma 6,  $a$  is thus at distance more than  $r + K$  of some hub center  $h \in H$ , so that  $R > 2(K + r)$  implies that  $d$  is at distance at least  $K + r$  from  $h$ . By definition of an hub-laminar decomposition,  $d$  is in a laminar  $L$ , which associated path is denoted by  $P = (v_0 = h, v_1 \dots v_r)$ .

Let  $v_{i_d}$  and  $v_{i_1}$  be vertices of  $P$  at distance less than  $k$  of  $d$  and  $x_1$ . As

$$d(d, y) \geq d(a, y) - d(d, a) \geq 3R + 4K + 1 - R - 1 \geq 2R + 4K, \quad (40)$$

$$d(d, f) \geq \min(d(d, y), l - 2R + 2) \geq 2R + 4K \quad (41)$$

Lemma 2 thus implies that every vertex at distance  $k$  of the path  $v_{i_d}, \dots, v_{i_d+2R+4K-2k}$  is  $K$  covered by  $P_{d,f}$ . So  $i_1$  is either lower than  $i_d$  or greater than  $i_d+2R+4K-2k$ .

But

$$d(v_{i_d}, v_{i_1}) \leq d(v_{i_d}, d) + d(d, x_1) + d(x_1, v_{i_1}) \leq 4K + 2k + 1 \quad (42)$$

$$d(h, v_{i_1}) \leq d(h, v_{i_d}) + 4K + 2k + 1 \leq d(h, v_{i_d}) + 2R + 4K - 2k \quad (43)$$

Thus,  $i_1 \leq i_d$ .

As  $B(v_{i_d+2R}, K)$  disconnects  $L$  and is entirely colored,  $Q$  is entirely in  $L$ . Let us denote  $v_{i_j}$  a vertex of  $P$  at distance  $k$  of  $x_j$  for every  $2 \leq j \leq t$ . Then for every  $1 \leq j < t$ ,  $d(v_{i_j}, v_{i_{j+1}}) \leq 2k + 1$ , so that all indices  $i_j$  are lower than  $i$  or all are greater than  $i + 2R + 4K - 2k$ . It follows that they are all lower than  $i_d$ .

But

$$i_d = d(h, v_i) \leq d(h, a) + d(a, d) + d(d, v_{i_d}) \leq 2R + 2k - r$$

so that  $d(h, v_{i_t}) \leq 2R + 2k - r$ .

Finally,

$$d(a, y) \leq d(a, h) + d(h, v_{i_t}) + d(v_{i_t}, y) \leq R - r + 2R + 2k - r + k < 3R + 4K + 1$$

which is a contradiction.

### Proof of Lemma 9

(i) Assume that  $h$  is at distance more than  $\max(R + r + 5K + 9k + 3, 3R + 4K)$  of every vertex  $a$  of  $A$ . By Lemma 8,  $h$  is of color  $lam$ . Therefore there is a path  $P_{d,f}$  and vertices  $a, a'$  in  $A$ ,  $u$  in  $P_{d,f}$  such that :

$$d(d, a) \leq R + 1 \quad (44)$$

$$d(f, a') \leq R + 1 \quad (45)$$

$$d(h, a) \geq R + r + 5K + 9k + 3 \quad (46)$$

$$d(h, a') \geq R + r + 5K + 9k + 3 \quad (47)$$

$$d(u, h) \leq K \quad (48)$$

By combining those inequalities :

$$d(d, u) \geq d(h, a) - d(d, a) - d(u, h) \geq r + 4K + 9k + 2 \quad (49)$$

$$d(f, u) \geq d(h, a') - d(d, a') - d(u, h) \geq r + 4K + 9k + 2 \quad (50)$$

By Lemma 4, there is  $vw \in E(G)$  and  $x$  at distance more than  $K + 6k$  from  $d$  and  $f$  such that  $d(x, v) = K$ ,  $d(P, w) = K + 1$ . It results that  $h$  should have been detected at line 20 and colored by a vertex of  $A$ .

Assume that  $h$  is at distance less than  $R + r + 5K + 9k + 3$  of a vertex  $a$  of  $A$  added at line 19. By lemma 6, there exist a hub of degree 1 of center  $h'$  such that  $d(a, h') \leq K + r$ . Therefore,  $d(h, h') \leq d(h, a) + d(a, h') \leq R + 2r + 6K + 9k + 3$ , which is impossible.

Let  $h$  be at distance less than  $R + r + 5K + 9k + 3$  of  $a$  added at line 11.  $a$  is at distance at most  $K + r$  of a hub  $h'$  of degree 3 or more. We then have :

$$d(h, h') \leq R + 2r + 6K + 9k + 3 < l \quad (51)$$

Thus  $h = h'$ , so that  $h$  is at distance at most  $K + r$  of  $a$ .

ii) Let  $h$  be an hub of degree 1. Suppose first that  $h$  is at distance less than  $3R + 4K$  of a vertex  $a$  of  $A$ . By a proof similar to (i),  $a$  was added at 19 and is at distance at most  $K + r$  of  $h$ .

Assume now that  $h$  is at distance more than  $3R + 4K$  of every vertex  $a$  of  $A$ . Then there is a path  $P_{d,f}$ , and  $u$  in  $P_{d,f}$  at distance at most  $K$  of  $h$ . We have,

$$d(d, u) \geq d(a, h) - d(a, d) - d(u, h) > 3k \quad (52)$$

$$d(f, u) > 3k \quad (53)$$

Let  $y$  and  $y'$  be the vertices of  $P_{d,f}$  at distance  $3k + 1$  of  $u$ . Let  $L$  be the laminar having  $h$  as an endpoint and  $P$  the path associated to  $L$ .

As  $3k + 1 < l$ ,  $y$  and  $y'$  are in  $L$ . Let  $v$  (resp.  $v'$ ) a vertex of  $P_L$  at distance at most  $k$  of  $y$  (resp.  $y'$ ). Assume w.l.o.g that  $v$  is closest to  $h$  than  $v'$ .

Then  $v \in P_{h,v'}$ , which leads to a contradiction by lemma 5. Thus,  $h$  cannot be at distance more than  $3R + 4K$  of every vertex  $a$  of  $A$ .

### Proof of Lemma 10

Denote by  $S$  the graph obtained by taking the union of all the paths of  $\mathcal{P}$ . We define  $S_1 = S \setminus \cup_{a \in A} B(a, 4K)$  and  $S_2 = S \setminus \cup_{a \in A} B(a, R)$ .

Let us first prove some intermediate claims concerning  $S_1$  and  $S_2$ .

**Claim 1:**  $S_1$  is a union of disjoint paths and every vertex of  $S_1$  is either in a laminar or in a hub of degree 2.

The only vertices of  $S$  of degree different from 2 are then exactly the vertices of  $H$  corresponding to hubcenters of degree different from 2. By Lemma 6, there exist a vertex  $a$  of  $A$  at distance at most  $2K$  of every such vertex  $h$ , implying  $B(h, 2K) \subset B(a, 4K)$ .

**Claim 2:**  $S_2$  corresponds to the set of deletable vertices of  $S$ . Moreover, every maximal path of  $S_2$  corresponds to a maximal path in  $S_1$ , which  $R - 4K$  first and last vertices have been deleted.

It is a direct consequence of  $R > 4K$ .

**Claim 3:** *Every deletable vertex is either in a hub of degree 2 or in  $B(S_1, k)$*

Vertices in a hub of degree  $\neq 2$  are undeletable by the same reasoning that for Claim 1. Consider now a vertex  $v$ ,  $k$ -covered by a vertex  $v'$  on a path of  $\mathcal{P}$ . If  $v$  is not  $k$ -covered by  $S_1$ , there exist  $a \in A$  with  $d(a, v') \leq 4K$  and thus  $d(a, v) \leq 4K + k < R$ .  $v$  is thus undeletable.

In  $G \setminus \cup_{a \in A} B(a, R)$ , we call *principal components* the components containing a path of  $S_2$ . By Lemma 3, a principal component contains no hub of degree three and is therefore either a portion of laminar or a succession of laminars and hubs of degree two, the first and last of them being incomplete. Moreover, every path  $Q$  built has to cross a principal component by Lemma 7.

Let us therefore consider a path  $Q$  joining  $a$  to  $a'$  and  $P_2$  a maximal path of  $S_2$  which principal component is crossed by  $Q$ . Let  $P_1$  be the maximal path corresponding to  $P_2$  and  $C$  the connected component of  $G \setminus \cup_{a \in A} B(a, 4K)$  that contains  $P_1$ .

$C$  is then composed of a succession of laminars or parts of laminars contained in  $B(S_1, k)$  and hubs of degree 2 and hubs of degree 2 hit by  $S_1$ .

**Claim 4:** *Every deletable vertex in  $C$  is deleted*

Let  $w$  be a deletable vertex in  $C$ .

**Case 1:**  *$w$  is in a hub of degree 2*

As the hub disconnects  $C$ ,  $Q$  hits the hub in some vertex  $u$ .  $w$  is then  $2r$ -covered by  $u$  and  $2r < K$ , so that  $w$  is deleted.

**Case 2:**  *$w \in B(S_1, k)$  but in no hub of degree 2*

$w$  is included in a laminar  $L$ , and is thus  $k$ -covered by a path  $P = p_0, \dots, p_s$  in  $\mathcal{P}$  which is included or partially included in  $S_1$ .

Note that if  $a$  (or  $a'$ ) has been added to  $A$  at line 15,  $a$  may be in  $L$ . For the sake of simplicity, we then still call  $L$  the part of  $L \setminus B(a, R)$  containing  $w$  and  $P$  the corresponding subpath of  $P$  starting at a vertex  $p_0$  at distance at most  $k$  from  $a$ . Such a vertex exists as  $a$  is then  $k$ -covered by  $P$ .

Let  $x$  and  $y$  be the first and last vertex of  $Q$  in  $L$ , and  $i_w, i_x$  and  $i_y$  be such that  $d(z, p_{i_z}) \leq k, \forall z \in \{w, x, y\}$ . We may suppose w.l.o.g. that  $i_x < i_y$ .

*Subcase 2.1:*  $i_x \leq i_w \leq i_z$

Then by Lemma 2,  $w$  is  $K$ -covered by  $Q$  and is thus deleted.

*Subcase 2.2:*  $i_w < i_x, p_0 \in H$

Then  $Q$  has to enter  $L$  by the hub  $B(p_0, r)$ , so that  $d(p_0, x) = r + 1$  and thus  $i_x \leq r + k + 1$ .

$$\begin{aligned} 2d(u, v) &\leq [d(x, p_{i_x}) + d(p_{i_x}, p_{i_w}) + d(p_{i_w}, w)] + [d(u, p_0) + d(p_{i_0}, p_{i_w}) + d(p_{i_w}, w)] \\ &\leq 3k + r + 1 + r + 1 + k \text{ as } d(p_0, p_{i_w}) + d(p_{i_w}, p_{i_x}) = i_x \\ &\leq 2R \\ d(u, v) &\leq R \end{aligned}$$

and  $v$  is thus deleted.

*Subcase 2.3:  $i_w < i_x, p_0 \notin H$*

This case corresponds to the case where  $a \in L$  has been added at line 15. Lemma 2 can then be applied between  $a$  and  $z$ , implying that  $w$  is  $K$ -covered by  $Q$  and thus deleted.

*Subcase 2.4:  $i_w > i_y$*

Symmetrical arguments of those used for Subcases 2.2 and 2.3 can be used.

Every deletable vertex is thus deleted in  $C$ . This implies that  $v$  can not be a deletable vertex of  $C$ . To prove the lemma, it is therefore sufficient to prove the following last claim.

**Claim 5:**  $B(Q, K + 1) \subset C \cup B(\{a, a'\}, R)$

Suppose it is not the case. W.l.o.g., suppose that there exist a vertex  $v$  neither in  $B(a, R)$  nor in  $C$  which is at distance  $R + 1$  of  $a$  and at most  $K + 1$  of  $q \in Q$ , the shortest path from  $v$  to  $q$  hitting  $B(a, R)$ . Let moreover denote by  $w$  the first vertex of  $Q$ .

**Case 1:**  $d(a, H) \leq 2K$

Then there exist  $h \in H$  with  $B(h, r) \subset B(a, R)$ . As  $\ell > 2R + K$ ,  $v$  and  $w$  are in laminars. Moreover, as  $d(a, h) \leq 2K < R - r - (K + 2)$ ,  $v$  and  $w$  are at distance at least  $K + 2$  of  $B(h, r)$ . By lemma 5,  $Q$  cannot hit  $w$  first and cross again  $B(h, r)$  to hit  $q$ .

Therefore,  $q$  is on the the subpath of  $Q$  between  $a$  and  $w$  and thus

$$\begin{aligned} d(a, v) + d(v, w) &\leq d(a, q) + d(q, v) + d(v, q) + d(q, w) \leq d(a, w) + 2K + 2 \\ R + 1 + d(v, w) &\leq R + 1 + 2K + 2 \\ d(v, w) &\leq 2K + 2 \end{aligned}$$

As  $v$  and  $w$  are in distinct laminars and at distance at least  $K + 2$  of  $B(h, r)$ , this is impossible.

**Case 2:**  $d(a, H) > 2K$

In that case,  $a$  has be added to  $A$  at line 15 and  $a$  belongs to a laminar  $L$  of path  $P = p_0, \dots, p_s$ . Let  $p_i$  be such that  $d(a, p_i) \leq k$ . Then  $2K - k \leq i \leq s - 2K + k$ . As  $K > 3k$ ,  $B(a, K)$  contains the subpath of  $P$  between  $p_{i-k}$  and  $p_{i+k}$  and all vertices  $k$ -dominated by it. This set disconnects  $L$  (any edge from one side to the other would lead to a chortcut for the path  $P$ ), ans so does  $B(a, K)$ .

The same reasoning that in Case 1 can then be done by replacing  $B(h, r)$  by  $B(a, K)$ .