Decomposing a Graph into Shortest Paths with Bounded Eccentricity

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

Abstract

We introduce the problem of hub-laminar decomposition which generalizes that of computing a shortest path with minimum eccentricity (MESP). Intuitively, it consists in decomposing a graph into several paths that collectively have small eccentricity and meet only at their extremities. We show that a graph having such a decomposition with long enough paths can be decomposed in polynomial time with bounds on the parameters of the decomposition. Moreover, such a decomposition with few paths allows to compute a compact representation of distances with additive distortion. The problem is related to computing an isometric cycle with minimum eccentricity (MEIC). We also show that having an isometric cycle with small eccentricity is related to the possibility of embedding the graph into a cycle with low distortion.

Keywords: Graph Decomposition, Graph Clustering, Distance Labeling, BFS, MESP

1. Introduction

The goal of this paper is to extend the MESP (Minimum Eccentricity Shortest Path) Problem from [Dragan and Leitert (2015)] and the related problem of recognizing $k$-laminar graphs from [Vökel et al. (2016)]. Both consist in finding a shortest path (in the sense that no path joining the same endpoints is shorter) $k$-dominating 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 (there is no longer shortest path in the graph). Relationships between the two parameters are derived in [Birmelé et al. (2016)].

To generalize this problem to more complex underlying structures, we introduce the problem of decomposing a graph into paths with bounded eccentricity. More precisely, we introduce the hub-laminar decomposition.
as a set of locally shortest paths that \(k\)-dominate 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 \(k\)-dominated by the path. Our definition requires that an edge between two nodes belonging to two different laminars must also belong to a hub. A laminar joins therefore two hubs, and laminars meet only inside hubs. The degree of a hub is then the number of laminars that meet in the hub. The main result of the paper is that computing such a decomposition becomes tractable when hub centers are far enough one from another, or equivalently when paths are long enough. We use two more parameters: the number \(\lambda\) of shortest path and the minimum length \(\ell\) of the paths. The MESP problem is equivalent to a hub-laminar decomposition with one laminar, i.e \(\lambda = 1\); and the \(k\)-laminar problem is when \(\lambda = 1\) and \(\ell\) is the graph diameter.

Such a generalization is naturally interesting in networks where one might want to identify a set of speedy linear routes that are “highly accessible” with applications in communication networks, transportation planning and water resource management. It is also motivated by DNA assembly in biology. DNA sequencing proceed through the reading of DNA fragments that must be assembled. When a single DNA strand is sequenced, comparison of fragments may lead to a graph with “laminar” structure (Völkel et al. (2016)), that is with large diameter and small shortest path eccentricity. In the context of metagenomics, several DNA strands are sequenced together and more complex structures appear (see Figure 1 in Völkel et al. (2016)). Identifying the laminar structures of such graphs is typically encountered in metagenomic approaches for evolution questions (see e.g. Saw (2015)). The problem of the assembly (gluing DNA fragments to reconstruct a DNA strand) is then mixed with that of binning (sorting DNA strands into groups that represent an individual genome or genomes from closely related organisms). See Thomas et al. (2012) for a presentation of assembly and binning problems in the context of metagenomics. Efficient decomposition of a graph into laminars could thus enhance the techniques for assembly and binning in this context.

The problem of decomposing a graph into \(\lambda\) laminars that \(k\)-dominate the graph is not well defined as there may be several trade-offs of parameters \(\lambda\) and \(k\). However, we show that when laminars are long enough compared to parameters \(r\) and \(k\), then all hub-laminar decompositions with these parameters are equivalent (same global structure) and have closely located hubs (except for hubs of degree two that do not affect the global structure). This
implies for example that the positions of the extremities of the minimum eccentricity shortest path (MESP) can be approximated within $O(k)$ distance when the diameter of a graph is large with respect to the eccentricity $k$ of the MESP. We define an algorithm that computes a hub-laminar decomposition under certain conditions. As the values of $r$ and $k$ are unknown, the algorithm is run with different values of parameters $R$ and $K$, the choice of those values will be discussed.

From a graph perspective, a very natural generalization of MESP is the problem of finding a minimum eccentricity isometric cycle (MEIC), that is a cycle preserving distances that has minimum eccentricity $k$. Note that such a cycle can be seen as a hub-laminar decomposition with two laminars and two hubs with degree two. An important motivation for the MESP problem is its relationship with embedding a graph into the line with small multiplicative distortion (Dragan and Leitert (2015)). We similarly show that the MEIC problem is related to embedding a graph into a circle with low multiplicative distortion, i.e. such that distances in the circle are within a constant factor of distances in the graph. Note that circle distortion is bounded by line distortion as a line segment can isometrically be embedded in a sufficiently long circle. (However, line distortion can be much larger than circle distortion.) Graph embedding in classical metrics is a well studied problem (Indyk (2001); Indyk and Matoušek (2004)). Another related subject with abundant literature is that of compactly representing the distances of a graph (Thorup and Zwick (2005); Peleg (2000)). We show that a decomposition with few laminars ensures a compact representation of distances with bounded additive distortion.

Related works: Finding a MESP is NP-complete but can be approximated within a constant factor (Dragan and Leitert (2015)). Better trade-off between computation time and approximation factor for MESP is obtained in Birmelé et al. (2016). The problem of efficiently representing the distances in a graph encompasses a vast literature dating from metric embedding (Asouad (1979)). Approximating embedding with low distortion is introduced in Badoiu et al. (2005a) where some results are provided in the case of the line. The case of embedding the metric induced by an unweighted graph is studied in Badoiu et al. (2005b). Embedding a graph metric into the line with minimum distortion is NP-complete but fixed parameter tractable with respect to distortion (Fellows et al. (2013)). Approximate distance oracles, i.e. compact data-structures for representing an approximation of distances,
are investigated in Thorup and Zwick (2005). A particular approach introduced by Peleg (2000) 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 Gavoille et al. (2004) and requires $\Omega(n)$ bits labels for general graphs. Approximation with a constant factor and sub-linear label size is derived in Thorup and Zwick (2005). Some results concern additive approximation such as Gavoille and Ly (2005) in the case of hyperbolic graphs. A longest isometric cycle can be found in polynomial time (Lokshtanov (2009)).

2. Definitions

We consider finite, undirected and connected graphs (the connectivity is always assumed within the paper). The vertex and edge sets of a graph $G$ are respectively denoted by $V(G)$ and $E(G)$. 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, or the set of edges between them. A path is a shortest path if its number of edges is minimal. For any vertices $u$ and $v$ on $P$, we denote by $P_{uv}$ the subpath of $P$ having $u$ and $v$ as endpoints.

We let $d_G(u,v)$ denote the distance between two vertices, i.e. the length of a shortest path from $u$ to $v$. When the graph $G$ is clear from the context, we omit the $G$ subscript and simply write $d(u,v)$. 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 set $B(U,r) = \bigcup_{u \in U} B(u,r)$. 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. $W \subseteq B(U,k)$. We say that $U$ has eccentricity $k$, denoted $ecc(U) = k$, when $k$ is the smallest integer such that $B(U,k) = V(G)$.

2.1. Hub-laminar decomposition

**Definition 1** (Hub-laminar decomposition). Consider a connected undirected graph $G$, two positive integers $r$ and $k$ with $k \leq r$, $H = \{h_1, \ldots, h_q\}$ a set of vertices of $G$ called hub centers, and $P = \{P_1, \ldots, P_p\}$ a set of paths of $G$ called laminar paths. A ball $B(h,r)$ with $h \in H$ is called a hub, and a
set \( B(P, k) \) with \( P \in \mathcal{P} \) is called a laminar. \((H, \mathcal{P})\) is an \((r, k)\)-hub-laminar decomposition of \( G \) if the following conditions are satisfied:

1. each laminar links two hubs centers: the endpoints \( h, h' \) of any \( P \in \mathcal{P} \) belong to \( H \) and for every other hub \( h'' \in H \setminus \{h, h'\} \),

\[
B(P, k) \cap B(h'', r + 1) = \emptyset
\]

2. laminars and hubs dominate \( G \): \( V(G) = \bigcup_{h \in H} B(h, r) \cup \bigcup_{P \in \mathcal{P}} B(P, k) \)

3. each laminar path is locally a shortest path: any path \( P \in \mathcal{P} \) with endpoints \( h \) and \( h' \) is a shortest path of \( G[B(P, k) \cup B(h, r) \cup B(h', r)] \)

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 \( u, v \in B(h, r) \).

The minimal laminar length of a decomposition \((H, \mathcal{P})\), denoted \( \ell \), is the minimal length of the paths in \( \mathcal{P} \). Its laminar size, denoted \( \lambda \), is the number of paths in \( \mathcal{P} \).

A hub-laminar decomposition \((H, \mathcal{P})\) with \( \ell \geq 2r + 1 \) forms a partition of the edges of \( G \) in the following sense: each edge is either inside exactly one hub (possibly touching many laminars ending in that hub), i.e. \( \exists h \in H \) s.t. \( u, v \in B(h, r) \); or, else, inside a unique laminar (possibly touching one hub extremity of that laminar), i.e. \( \exists ! P \in \mathcal{P} \) s.t. \( u, v \in B(P, k) \).

Figure 1 illustrates this definition and the notion of quotient graph that we define next. This definition basically defines a decomposition into \( \lambda k \)-neighborhoods of internally far apart shortest paths. It may seem a bit involved, but we think it expresses in a minimalist way what we mean by “internally far apart” with Axiom 4. Axioms 1 and 2 indicate that the graph is decomposed into laminars which are \( k \)-neighborhoods of certain paths and hubs which are balls centered at the extremities of those paths. Axiom 3 requires a path to be shortest in the induced graph (rather than in \( G \)), to allow laminars with different length between the same two hub centers.

2.2. Quotient graph and equivalence between decompositions

As previously mentioned, the hub-laminar decomposition gives naturally raise to a skeleton, which can be simplified into a quotient graph.
Figure 1: Illustration of an hub-laminar decomposition with \( r = 2, k = 1 \). 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).

**Definition 2** (quotient graph and reduced quotient). Given a graph \( G \) and an \((r, k)\)-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 \( P \in \mathcal{P} \) with endpoints \( h, h' \) there is an edge \( hh' \) whose label is the length of \( P \).

The degree of a hub denotes the degree of the corresponding vertex in the quotient graph, or equivalently the number of laminar paths its center is the endpoint of.

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

When the quotient is not a cycle (a case specifically adressed by MEIC, see Section [3]) the reduced quotient is well defined and unique (recall that graphs are supposed connected).

**Definition 3** (equivalence between decompositions). Two hub-laminar decomposition of a given graph \( G \), possibly with different parameters \( r, k \), are \( D \)-equivalent if they have the same non edge-labeled reduced quotient graph, up to an isomorphism \( \phi \) of vertex-sets such that \( d_G(h, \phi(h)) \leq D \).
2.3. Isometric cycle, circle embedding and distance labeling

A cycle $C$ in a graph $G$ is isometric if it preserves distances, i.e. $d_C(u, v) = d(u, v)$ for all $u, v \in V(C)$. In other words, for any pair $u, v$ of nodes on the cycle, one of the two paths linking $u$ and $v$ in the cycle is a shortest path in the graph. Note that an isometric cycle is necessarily an induced cycle. The MEIC problem consists in finding an isometric cycle with minimum eccentricity.

It can be shown to be NP-complete following a proof similar to one used to show the NP-completeness of the MESP problem. Indeed, [Völkel et al. 2016] shows that the MESP problem is NP-complete by exhibiting a family of graphs such that computing the MESP between two vertices called $V_1$ and $V_{2n}$ is solving an associated 3SAT formula. By using the same family of graphs and adding an edge between the vertices $V_1$ and $V_{2n}$ we get a reduction from 3SAT to the MEIC problem.

A circle embedding of a graph $G$ is a mapping $f : V(G) \rightarrow C$ where $C$ is a circle of given length $c$. It has distortion $\gamma$ if $d_G(u, v) \leq d_C(f(u), f(v)) \leq \gamma d_G(u, v)$ for all $u, v \in V(G)$. The circle distortion $cd(G)$ of $G$ is the minimum distortion of a circle embedding of $G$.

A distance labeling of a graph $G$ consists in assigning a label $L_u$ to each node $u \in V(G)$ together with a distance estimation function $f$ that outputs an estimation of $d(u, v)$ when given $L_u$ and $L_v$ as input. It has additive distortion $\alpha$ if $d(u, v) \leq f(L_u, L_v) \leq d(u, v) + \alpha$ for all $u, v$ in $G$.

3. Main results

Obviously, the reduced quotient graph of a graph having a $(r, k)$-hub-laminar decomposition follows the following trichotomy: it is either a path, a cycle or has a node of degree at least three.

We treat separately the three cases. In the first case, the graph has a shortest path with eccentricity at most $\max \{3k; 2r\}$ and can be recognized through an approximate MESP algorithm such as Birmelé et al. (2016). (The bound $\max \{3k; 2r\}$ is a consequence of Lemma 3 given in Section 4.) In the second case, the graph has an isometric cycle with eccentricity at most $\max \{3k, 2r\}$. To recognize such graphs, we propose an approximation MEIC algorithm:
Theorem 1. Given a graph containing a \( k \)-dominating isometric cycle with length \( \ell \), a \( 6k \)-dominating isometric cycle can be computed in \( O(n^{4.752} \log(n)) \) time. Moreover, the computed cycle is \( 3k \)-dominating when \( \ell > 12k + 2 \).

The proof of this theorem may be found in Section 4.1. We obtain therefore an algorithm for approximating circle embedding with low distortion.

Proposition 1. If a graph has circle distortion \( \gamma \), it is possible to embed it into a circle with distortion \( O(\gamma^2) \) in polynomial time.

This proposition follows from Theorem 1, Proposition 3, and Proposition 4 in Section 6.1.

Recognizing the general case of decomposition is not a well defined problem as several decompositions may yield different trade-offs of the parameters. For the same graph both a \((r,k)\)-hublaminar decomposition and a \((r',k')\)-hublaminar decomposition may exist and have completely different shapes. However, when laminars are long enough, all \((r,k)\)-hub-laminar decompositions are indeed \( O(k) \) equivalent. This can be seen as a consequence of the following recognition result, our main theorem which proof stretches over Section 4.

Theorem 2. Given a graph \( G \) having a \((r,k)\)-hub-laminar decomposition \((H, \mathcal{P})\) of minimal laminar length \( \ell \geq 10r + 52k + 5 \) and integers \( K, R \) such that \( K \geq 3k, R \geq 4K + 3r \) and \( 2R + 8K < \ell - 4r - 4k - 4 \), it is possible to compute in \( O(\min(n,\lambda)m) \) time a \((K,R)\)-hub-laminar decomposition which is \((K + 2r + k)\)-equivalent to \((H, \mathcal{P})\).

From the graph metric point of view, we obtain then a compact representation of distances:

Proposition 2. Given a graph \( G \) having an \((r,k)\)-hub-laminar decomposition with laminar size \( \lambda \), it is possible to compute in polynomial time a \( O(\max\{k,r\})\)-additive distance labeling with \( O(\lambda \log n) \) bit labels.

This is proven as Proposition 5 in Section 6.2

4. Algorithms

4.1. Minimum Eccentricity Isometric Cycle (MEIC) Problem

We propose to approximate the MEIC Problem by computing a longest isometric cycle, that is an isometric cycle of \( G \) with maximum length, since
such a cycle $O(k)$-dominates any $k$-dominating isometric cycle (Lemma 2).

For any cycle $C$ and any pair of vertices $a$ and $b$, we denote by $C_{a,b}$ and $C_{b,a}$ the two paths in $C$ linking $a$ and $b$.

**Lemma 1.** Let $G$ be a graph with an isometric cycle $C$ $k$-dominating $G$. Let $u$ and $v$ be any two vertices, and $u'$ and $v'$ be two vertices on $C$ that are at distance at most $k$ of respectively $u$ and $v$.

Every path between $u$ and $v$ $2k$-dominates either $C_{u',v'}$ or $C_{v',u'}$.

**Proof.** Let $P$ be a path between $u$ and $v$. Suppose that $P$ does not $2k$-dominate some vertex $b$ on the path $C_{v',u'}$ and consider any vertex $a$ in $C_{u',v'}$.

Without loss of generality, assume that $u'$ (resp. $v'$) is in the path $B = C_{a,b}$.

Then $u$ is at distance at most $k$ of $C_{a,b}$ and $v$ is at distance at most $k$ of $C_{b,a}$. Moreover, as every vertex of $G$ is at distance at most $k$ of one of those two paths, there exist $c$ and $d$ that are adjacent vertices in $P$ such that $c$ is at distance at most $k$ of $C_{a,b}$ and $d$ at distance at most $k$ of $C_{b,a}$.

As $d(c',d') \leq d(c',c) + d(c,d) + d(d,d') \leq 2k + 1$ and $C$ is an isometric cycle, either $C_{c',d'}$ or $C_{d',c'}$ is of length at most $2k+1$ and is thus $2k$-dominated by $\{c,d\}$. Furthermore $b$ and $a$ are not in the same subpath of $C$ between $c'$ and $d'$, hence either $a$ or $b$ is $2k$-dominated by $\{c,d\}$. As $b$ cannot be $2k$-dominated by $P$ it follows that $a$ is $2k$-dominated by $\{c,d\}$ hence by $P$.

The previous claim being true for every $a$ in $C_{u',v'}$, the lemma follows.

**Lemma 2.** Let $G$ be a graph with an isometric cycle $C$ $k$-dominating $G$. Let $D$ be a longest isometric cycle of $G$.

Every vertex of $C$ is then at distance at most $4k$ of $D$. Furthermore, if $D$ is of length more than $8k + 2$ then every vertex of $C$ is at distance at most $2k$ of $D$.

**Proof.** Let $C = c_1,...c_p$ and assume that $D$ does not $2k$-dominate $C$. Without loss of generality, we may assume that $c_1$ is at distance greater than $2k + 1$ of every vertex of $D$.

Let $c_i$ and $c_j$ be vertices at distance less than $k$ of $D$ and such that $C_{c_1,c_i-1}$ and $C_{c_{j-1},c_1}$ contain no vertex at distance less than $k$ of $D$.

Let us note $D = d_1,...d_q$, and define a function $f$ from $[1,q+1]$ to $[1,p]$ such that for every $x$ in $[1,q]$, $c_{f(x)}$ is at distance at most $k$ from $d_x$. 


and such that \( f(q+1) = f(1) \). We may assume, w.l.o.g., that \( c_f(1) = c_i \), that is \( f(1) = i \). Note that, for every \( x \in [1, q[, \)

\[
i = f(1) \leq f\left(\left\lfloor \frac{q}{2} \right\rfloor\right) \leq j
\]

It is then sufficient to show that there exist \( x \) in \([1, \left\lfloor \frac{q}{2} \right\rfloor]\) such that :

\[
|f(x) - f\left(\left\lfloor \frac{q}{2} \right\rfloor + x\right)| \leq 2k + 1
\]

In other words, that there exists two opposite vertices in \( D \) at distance at most \( 4k + 1 \), which implies \( |D| \leq 8k + 2 \).

If \( f\left(\left\lfloor \frac{q}{2} + 1\right\rfloor\right) - f(1) \leq 2k + 1 \), this result is straightforward. If not, let \( x \) be a value in \([1, \left\lceil \frac{q}{2} \right\rceil]\) such that :

\[
f(x) \leq f\left(\left\lfloor \frac{q}{2} \right\rfloor + x\right)
\]

\[
f(x + 1) \geq f\left(\left\lfloor \frac{q}{2} \right\rfloor + (x + 1)\right)
\]

Such an \( x \) exists as the first equality holds for \( x \) equals to 1 and the second equality holds for \( x \) equals to \( \left\lceil \frac{q}{2} \right\rceil \).

Assume that \( f\left(\left\lfloor \frac{q}{2} \right\rfloor + x\right) - f(x) > 2k + 1 \), as otherwise the result is again straightforward.

\[
|f(x + 1) - f(x)| \leq d(c_f(x), d_x) + d(d_x, d_{x+1}) + d(d_{x+1}, c_f(x+1)) \leq 2k + 1
\]

then implies

\[
f\left(\left\lfloor \frac{q}{2} \right\rfloor + (x + 1)\right) \geq f\left(\left\lfloor \frac{q}{2} \right\rfloor + x\right) - (2k + 1) > f(x) \geq f(x + 1) - (2k + 1)
\]

We get, combining the inequalities, that

\[
f(x + 1) \geq f\left(\left\lfloor \frac{q}{2} \right\rfloor + x + 1\right) > f(x + 1) - 2k - 1
\]

and thus

\[
|f(x) - f\left(\left\lfloor \frac{q}{2} \right\rfloor + x\right)| \leq 2k + 1
\]
$D$ is therefore of length at most $8k + 2$ if it does not $2k$-dominate $C$, which proves the second statement of the Lemma.

To prove the first one, it is now sufficient assume that $D$ is of length $p \leq 8k + 2$ and prove it $4k$-dominates $C$. To do so, consider two opposite vertices $u$ and $v$ on $D$, that is at distance at least $\lfloor \frac{p}{2} \rfloor$.

Let $c_i$ (resp. $c_j$) in $C$ at distance less than $k$ of $u$ (resp. $v$). Then,

$$d(c_i, c_j) \geq d(u, v) - d(c_i, u) - d(v, c_j) \geq \left\lfloor \frac{p}{2} \right\rfloor - 2k$$

As $D$ is a longest isometric cycle and thus $|C| \leq p$,

$$|C_{c_j, c_i}| \leq |C| - d(c_i, c_j) \leq \left\lfloor \frac{p}{2} \right\rfloor + 2k \leq 6k + 1$$

Similarly, $|C_{c_i, c_j}| \leq 6k + 1$. Hence, for every $c_i$ in $C$, $d(c_i, c_i) \leq 3k$ or $d(c_j, c_i) \leq 3k$.

As $u$ (resp. $v$) is at distance $k$ of $c_i$ (resp. $c_j$), $d(u, c_i) \leq d(u, c_i) + d(c_i, c_i) \leq 4k$ or $d(v, c_i) \leq 4k$

Consequently, a longest isometric cycle in a graph is a 5-approximation for the MEIC problem, and a 3-approximation when the graph has a large enough diameter. As shown in [Lokshtanov (2009)], a longest isometric cycle can be computed in $O(n^{4.752} \log(n))$ time. Theorem 1 is thus a direct consequence of this and Lemma 2. The bound for the 3-approximation when the graph has a large enough diameter is tight, for the 5-approximation, we have found an instance that shows that it is at best a 4-approximation.

### 4.2. General case outline

The previous subsection corresponds to the case where the quotient graph is a cycle. The case where it is a path is solved by [Birmelé et al. (2016)]. These two cases cover all situations where the quotient graph has maximum degree at most 2. Consider now a graph $G$ having a $(r,k)$-hub-laminar decomposition $(H, \mathcal{P})$ of minimal laminar length $\ell$ and having at least one hub of degree at least 3. Notice that in the sequel, we always refer to $H$, $\mathcal{P}$ or the parameters $r, k, \lambda$ and $\ell$ but of course we do not know them and they are not part of the input.
We first present the algorithm and state the theoretical results related to each step. Technical proofs and intermediate lemmas are postponed to Section 5 to preserve readability.

In all that section, the assumptions of Theorem 2 are considered true: we are given input parameters $K$ and $R$ satisfying $K \geq 3k$, $R \geq 4K + 3r$ and $\ell > 2R + 8K + 4r + 4k + 4$.

The underlying idea of the algorithm is to use BFS (Breadth-first search) to compute shortest paths and their $K$-neighborhoods, in order to use the following central lemma from Birmelé et al. (2016). It states that any path going through a laminar $3k$-dominates the central part of it, that is all vertices not to far from the two corresponding hub centers.

**Lemma 3** (Path local domination). Consider a laminar path $P \in \mathcal{P}$. Let $Q$ be a path from $u$ to $v$ contained in $B(P, k)$. Let $u' \in P$ and $v' \in P$ 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 $Q$. Furthermore, every vertex of $B(P_{u'v'}, k)$ is at distance at most $3k$ from $Q$.

By that lemma, the choice $K \geq 3k$ will ensure the domination of every laminar traversed by well-chosen shortest paths. However, a set of vertices approximating the set $H$ has first to be chosen according to the following definition.

**Definition 4.** A vertex $a$ dominates a hub-center $h \in H$ if $d(a, h) \leq K + 2r + k$.

A vertex set $A$ is $H$-close if every vertex of $A$ dominates a vertex of $H$, no vertex of $H$ being dominated by two vertices of $A$.

A vertex set $A$ is $H$-dominating if it is $H$-close and every vertex of $H$ defining a hub of degree different from 2 is dominated by a vertex of $A$.

The first part of the algorithm, called FindHubs and detailed in Section 4.4, determines a $H$-dominating set $A$ containing the hub-centers of the returned decomposition.

Note that $\ell > 2(K + 2r + k)$ implies that no vertex of $A$ can dominate two different vertices of $H$. Therefore, an $H$-dominating set $A$ is an approximation of $H$ in the sense that $A$ contains exactly one vertex dominating the center of every hub of degree 1, 3 or more, even if it may contain or not a vertex dominating the center of every hub of degree 2. The special status of hubs of degree 2 is due to the fact that they may be integrally included in
the $K$-neighborhood of shortest paths if $K > r$, so that they may be difficult to detect. Although the returned quotient graph may differ from the original quotient graph in such cases, the reduced quotient graphs will coincide.

The former point raises a particular difficulty in configurations corresponding to a cycle in the quotient graph of $(H, \mathcal{P})$ containing only one hub of degree at least 3, like the two laminars on the right of Figure 2. This is called a Problematic Configuration: there exists at least a degree 2 hub $h \in H$ somewhere in that cycle, but one might not be able to detect it. In that case, a vertex $b$ situated in the middle of the cycle is added to a set $B$ which will be returned together with $A$ by $\text{FindHubs}$.

![Figure 2: The decomposition $(H, \mathcal{P})$ is unknown. A problematic configuration is on the right of the graph: a laminar cycle with only one hub having degree different from 2.](image)

The laminars are determined in a second step by the $\text{FindLaminars}$ procedure, which basically links the hub-centers of the previous step by shortest paths. One technical point has to be taken into account: a hub of degree 2 may have been missed by $\text{FindHubs}$ and can be discovered at that step. In that case, the set of hubs $A$ is adapted by adding the new discovered hub, and if needed, the corresponding vertex is deleted from $B$.

Figure 2, 3 and 4 give a summary of the two steps by showing a possible outcome of the $\text{FindHubs}$ and $\text{FindLaminars}$ on an example.
Figure 3: During an execution of the function \textit{FirstHub}, a set $A$ of hub centers is computed such that every ball $B(h, r)$ for $h \in H$ having degree different from 2 is covered by $B(a_i, R)$ for some $a_i \in A$. Tentative hubs such as $b_1$ may be added in the middle of problematic configurations and returned in a set $B$.

4.3. Topology of $G \setminus B(A, R)$

Both procedures \textit{FindHubs} and \textit{FindLaminars} will rely on BFS trees covering connected components of $G \setminus B(A, R)$, $A$ being an $H$-close set. Before detailing those procedures, the following lemma explicits the possible topologies of such components with respect to the decomposition $(H, \mathcal{P})$. Figures 5, 6 and 7 illustrate all possibilities.

Lemma 4.

Let $A$ be an $H$-close set and $g$ a connected component of $G \setminus B(A, R)$. $g$ has one of the mutually exclusive following topologies:

Type a) $g$ contains no hub and touches only one set $B(a, R)$, $a \in A$;

Type b) $g$ contains a hub of degree at least three;

Type c): there exist a sequence of hubs and laminars $H_0, L_1, H_1, \ldots, L_z, H_z$, $z \geq 1$, such that the center $h_0$ of $H_0$ is dominated by some node $a_1 \in A$, $H_z$ is of degree 1, all other hubs (if $z \geq 2$) are of degree two, and $g$ is consisted exactly of the union of the vertices in these hubs and laminars except those in $B(a_1, R)$. 

14
Figure 4: When computing \textit{FindLaminars}, tentative hubs in set $B$ might be modified: in this example, $b_1$ becomes $a_6$. Some thin hubs of degree 2 are not detected and belong to $K$-laminars, as illustrated by the top-left hub for instance. In any case, the reduced quotient graph stays the same even if some hubs of degree 2 are missed.

\textbf{Type d):} there exist a sequence of hubs and laminars $H_0, L_1, H_1, \ldots, L_z, H_z$, $z \geq 1$ and $H_z \neq H_0$, such that the centers of $H_0$ and $H_z$ are dominated, all other hubs (if $z \geq 2$) are of degree two, and $g$ is composed of all the vertices in those sets except those in $B(A,R)$;

\textbf{Type e) (problematic configuration):} there exist a sequence of hubs and laminars $H_0, L_1, H_1, \ldots, L_z, H_0$, $z \geq 1$, such that the center of $H_0$ is dominated by some node $a_1 \in A$, all other hubs are of degree two, and $g$ is composed of all the vertices in those sets except those in $B(a_1,R)$.

Moreover, every vertex neighboring $B(a,R)$, $a \in A$, belongs to a laminar incident to $B(h,r)$, $h \in H$ being the hub center dominated by $a$.

\subsection*{4.4. Finding hubs}

The hub detection algorithm \textit{FindHubs} relies on a vertex-coloring procedure of $G$, which is initially completely uncolored. The vertices are then colored gradually by the procedure \textit{NextHub}, and some of them added to sets $A$ or $B$ in a way such that the following Invariants are satisfied at each step:
Figure 5: Illustration of a connected component of Type a). The laminar path is in bold, $k = 2$, $R = 3$ and the squared vertices are in $B(a, R)$ colored in light grey. The connected component of Type a), in dark grey, contains 3 vertices. All 3 are at distance more than $R$ of $a$ and at distance at most $k$ of the laminar path. However $B(a, R)$ disconnects them from the rest of the laminar.

Figure 6: Illustration of a connected component of Type b) (on the left) and a component of Type c) (on the right). Vertices $a_1$, $a_2$ and $a_3$ correspond to vertices of $A$ already detected. $B(A, R)$ is colored in light grey. The connected components are in dark grey.

**Invariant 1:** All balls $B(a, R)$ for $a \in A \cup B$ are disjoint and for each $a \in A$, all nodes in $B(a, R)$ are colored with a color specific to $a$.

**Invariant 2:** Some connected components of $G \setminus B(A, R)$ may be colored with a specific color $lam$ (as laminar).

**Invariant 3:** The set $U$ of uncolored vertices is a union of connected components in $G \setminus B(A, R)$.

**Invariant 4:** $A$ is $H$-close.

**Invariant 5:** Every colored $h \in H$ defining a hub of degree 1 or at least 3 is dominated by a vertex $a \in A$.

To start, $FindHubs$ needs a first vertex $a \in A$ dominating some $h \in H$. Coloring the vertices of $B(a, R)$ then ensures that the five Invariants are sat-
Figure 7: Illustration of a connected component of Type d) (on the left) and a component of Type e) (on the right). Vertices \(a_1\), \(a_2\) and \(a_3\) correspond to vertices of \(A\) already detected. \(B(A,R)\) is colored in light grey. The connected components are in dark grey.

Finding \(H\)-Hubs then consists in applying as long as possible the \(NextHub\) procedure, introduced in Section 4.4.2 which colors a non-empty set of uncolored vertices by preserving the five Invariants. \(NextHub\) is repeated until no vertex of \(A\) has an uncolored vertex at distance \(R + 1\). Invariant 1 and 3 then ensure that the whole graph is colored, Invariants 4 and 5 then implying that \(A\) is \(H\)-dominating.

The initialisation of the procedure is postponed to section 4.4.3.

\(NextHub\) relies on several BFSs which are run from the border of \(B(A,R)\) in the connected components of \(G \setminus B(A,R)\). As shown later on, Type a) and Type e) components are then characterized by the fact that the furthest node in the BFS is close to the ball \(B(a,R)\) it started from. Therefore, the uncolored nodes close to \(a\) need to be marked to test whether we stop near \(a\) or not (as we shall see near \(a\) means in \(B(a,R + 2K + 1)\)). The marks are removed after the BFS is done.

4.4.1. The StopBFS function

Provided a marked and uncolored vertex \(d\) and a color \(c\), the \(StopBFS\) procedure consists in running an usual Breadth-first search starting at vertex \(d\), with the following additional rules:

- only uncolored or marked vertices are put in the BFS queue.
- if a vertex is visited (i.e. extracted from BFS queue) and has a colored neighbor whose color is not \(c\) then this (uncolored) vertex is noted \(f\) and the rest of the BFS is computed.
- if the BFS queue becomes empty without encountering the previous case, let \(f\) be its deepest leaf such that there exists an unmarked vertex
on the BFS path from \(d\) to \(f\); if such a leaf doesn’t exist, i.e. only marked vertices were traversed, let \(f = d\).

- the function \(\text{StopBFS}(d, c)\) returns the BFS tree \(T\) as well as the path \(Q\) from \(d\) to \(f\) in that tree.

\(\text{StopBFS}\) will be applied at the first step of \(\text{NextHub}\). Note that Invariant 3 then implies that the explored vertices correspond to a connected component of \(G \setminus B(A, R)\), or to a subgraph of such a component if another color is encountered.

4.4.2. Finding a new hub: \(\text{NextHub}\)

The procedure \(\text{NextHub}\) can now be described. It relies on the following result, which allows to detect hubs, that is to select vertices in \(A\) which are close to vertices in \(H\).

**Lemma 5** (Hub trigger). Let \(Q\) be a shortest path returned by \(\text{StopBFS}\), or any shortest path in \(G\). Denote by \(r_{3K}(Q)\) the subpath of \(Q\) obtained by removing the \(3K\) first and \(3K\) last vertices along that path.

Suppose there exist a vertex \(u \in r_{3K}(Q)\) and an edge \(vw \in E(G)\) such that \(d(u, v) = K\) and \(d(Q, w) = K + 1\). Then there exist a hub center \(h \in H\) dominated by \(u\).

Conversely, suppose that the set of vertices explored by \(\text{StopBFS}\) contains \(B(h, \frac{R}{2} - R)\) for some \(h \in H\) defining a hub of degree at least \(3\), the vertices \(d\) and \(f\) of \(\text{StopBFS}\) being outside this set. Suppose moreover that the path \(Q\) output by the \(\text{StopBFS}\) intersects \(B(h, r)\). Then there exist a vertex \(u \in r_{3K}(Q)\) and an edge \(vw \in E(G)\) such that \(d(u, v) = K\) and \(d(Q, w) = K + 1\).

The following lemma lists the behaviors of the result of \(\text{StopBFS}\) depending on the explored subgraph, leading thus to an algorithm based on the result of the \(\text{StopBFS}\) procedure.

**Lemma 6.** Suppose that Invariants 1 to 5 are fulfilled and consider \(a \in A\). Assume that the set of marked vertices is \(B(a, R + 2K + 1)\) and that \(\text{StopBFS}\) is run from a vertex \(d\) such that \(d(a, d) = R + 1\). Let \(g\) be the connected component of \(G \setminus B(A, R)\) explored (partially) by \(\text{StopBFS}\), and denote by \(Q\) and \(f\) the path returned by \(\text{StopBFS}\) and its last vertex.

Depending on the topology type of \(g\) as defined in Lemma 4, the following holds:
Type a): all explored vertices are marked, and thus $f = d$;

Type b): there exist a triple of vertices satisfying the conditions of Lemma 5;

Type c): $f$ and its neighborhood are unmarked and uncolored. In that case, $f$ dominates the center of $H_z$;

Type d): $f$ is unmarked and neighbors a colored vertex of a ball $B(a', R)$ centered at a vertex $a' \in A \setminus \{a\}$ that dominates the center of $H_z$;

Type e): $f$ is marked and different from $d$.

The NextHub procedure, whose pseudo-code is given by Algorithm 1, consists in determining which of the five cases is relevant by testing the presence of a triple satisfying Lemma 5 or looking at the status of $f$. Note that a triple satisfying Lemma 5 may also be found in components of Type c), d) or e), corresponding then to the detection of a hub of degree 2.

In each case, either a vertex dominating some uncolored $h \in H$ can be added to $A$, or a whole component of $G \setminus B(A, R)$ containing no hub or only hubs of degree two can be colored. In the last case, corresponding to the problematic configuration, a temporary hub $b \in B$ is arbitrarily added in the middle of $Q$.

Assuming a correct initialisation, the following lemma implies the validity of FindHubs.
**Algorithm 1:** Pseudo-code of the NextHub procedure

```
NextHub

Input: A graph G with possibly colored vertices, integers R and K, hub-center sets A and B (B is possibly empty), and a vertex a ∈ A

Output: Updated sets A, B and vertex coloring

Mark every uncolored vertex in B(a, R + 2K + 1)

Choose a marked vertex d at distance R + 1 from a

Let (T, Q) = stopBFS(d, col(s)) and f be the last vertex of Q

If f = d then
   /* Case a) */
   Color all vertices of T with color lam

else if ∃w, u s.t. w is uncolored and u ∈ r_{3K}(Q) and d(w, u) = K + 1 and d(w, Q) = K + 1 then
   /* A triple satisfying Lemma 5 is found. */
   Add to A the first vertex u of r_{3K}(Q) satisfying the above condition.
   Color every vertex in B(u, R) with a new color col(u)

else if f is a marked vertex then
   /* Case e) */
   Add to B the vertex b in the middle of Q
   Color all vertices of T with color lam

else if f is an uncolored vertex and has a colored neighbor then
   /* Case d) */
   Color all vertices of T with color lam

else
   /* Case c) */
   Add f to A
   Color every vertex in B(f, R) with a new color col(f)

Unmark every marked vertex
```

**Lemma 7.** If NextHub is run on a colored graph G and a set A such that the Invariants 1 to 5 are verified, the modified graph and set obtained as outputs also satisfy them.

**Proof.** Invariants 1 and 2 are straightforward given the algorithm.
Invariant 3 is conserved as in every configuration, either a vertex is added to $A$ and its $R$-neighborhood is colored, or a whole connected component of $G \setminus B(A, R)$ is colored with color $lam$.

The conservation of Invariants 4 and 5 is a consequence of Lemma 5 if a triple satisfying it is found and of Lemma 6 in case c). The three other cases do not change the set $A$ nor color hubs of degree different from 2.

4.4.3. Initialisation

In order to use the $NextHub$ procedure while conserving the Invariants, a first vertex has to be added to $A$. This is done by using $NextHub$ as follows.

A first BFS is performed starting at a vertex $s$ chosen arbitrarily. Let $x$ be a furthest vertex from $s$ and let $Q$ be the shortest path between $s$ and $x$ computed by the BFS. If $Q$ contains a triplet of vertices $(u, v, w)$ as defined by Lemma 5 then $u$ is chosen as the first hub. Otherwise, let $m$ be a vertex in the middle of $Q$ and let $d$ be the vertex of $Q$ at distance $R + 1$ of $m$ and the closest to $s$.

The $NextHub$ procedure is then called with $A = \{m\}$, the vertices of $B(m, R)$ are colored and the $stopBFS$ starts on $d$. The resulting procedure $FirstHub$ for finding a first hub center is detailed as Algorithm 2. A vertex $u$ dominating a vertex of $H$ is then detected as shown with the following lemma.

**Lemma 8.** Assume that $(H, P)$ contains at least one hub of degree at least 3. The procedure $FirstHub$ described above returns a vertex $u$ corresponding to the configuration of Lemma 5.

4.5. Finding laminars

Given the $H$-dominating set $A$ and the set $B$ pointing to the problematic configurations, the procedure $FindLaminars$ constructs shortest paths between vertices of $A$, which will be the laminar paths of the returned decomposition.

Again, this is done by applying repeatedly BFS, but rooted on vertices of $A$. For each constructed path $Q$ linking two hub centers $a$ and $a'$, the vertices of the corresponding constructed laminar $B(Q, K)$ are removed from the graph, except those of the hubs $B(a, R)$ and $B(a', R)$ which are declared undeletable. The process ends when the graph consists in disconnected hubs only.
Algorithm 2: Pseudocode of function FirstHub

Two technical difficulties however have to be taken into account. The first one is the possible discovery of hubs of degree 2 which had been missed by FindHubs; it can easily be handled by updating $A$. The second one resides in problematic configurations as a laminar has to link two distinct hubs. In order to solve it, vertices of $B$ are considered first. More precisely, for every $b \in B$, there exist $a \in A$ such that, for any BFS traversal starting from $b$, the first encountered element of $A \cup B$ is $a$. Thus, two BFS from $b$ to $a$ are run, following the Type e) component in opposite directions. Either a hub of degree 2 is discovered, $A$ is updated and $b$ can be discarded, or the two obtained paths $K$-dominate all vertices of the Type e) component and $b$ is transferred to $A$.

The pseudo-code of FindLaminars is given in Algorithm 3.
FindLaminars

Input: A graph $G$, integers $R$ and $K$

Output: a hub-laminar decomposition $(A, Q)$

$(A, B) = \text{FindHubs}(G, R, K)$

$Q = \emptyset$

Mark all vertices as deletable

For each $a \in A$ do

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

For each $b \in B$ do

Run a BFS starting at $b$ and stopping on the first vertex $a \in A$

Let $Q_1$ be the path from $b$ to $a$ computed by this BFS

Run a BFS starting at $b$, not using vertices of $B(Q_1, K) \setminus (B(b, R) \cup B(a, R))$ and stopping in $a$

Let $Q_2$ be the path from $b$ to $a$ computed by this BFS

Compute $g$, the union of $B(a, R)$ and of the connected component of $G \setminus B(a, R)$ containing $b$

Color in $g$ the vertices of $B(a, R)$, $B(b, R)$, $B(Q_1, K)$, $B(Q_2, K)$

If $\exists$ an uncolored vertex $c$ in $g$ then

Add $c$ to $A$

Mark the vertices in $B(c, R)$ as undeletable

else

Add $b$ to $A$

Mark the vertices in $B(b, R)$ as undeletable

Delete from $G$ the deletable vertices of $B(Q_1, K) \cup B(Q_2, K)$

Add $Q_1$ and $Q_2$ to $Q$

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

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

Let $Q$ be the path from $a$ to $a'$ computed by this BFS

If $\exists w, u$ s.t. $h \in r_{3K}(Q)$, $d(w, u) = K + 1$, $d(w, Q) = K + 1$ then

Add to $A$ the first vertex $h$ of $Q$ satisfying the above

Mark the vertices in $B(h, R)$ as undeletable

else

Add to $Q$ the path $Q$ from $a$ to $a'$ computed by this BFS

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

Algorithm 3: Pseudo-code of the FindLaminars procedure
Lemma 9. Suppose that FindLaminars is run, with the sets $A$ and $B$ returned by FindHubs as its input. Let us consider the evolution of the deletable vertices during the algorithm. The following holds:

1. After each iteration of the For or the While loop, the set of deletable vertices is a union of components of $G \setminus B(A, R)$ of Type a), d) or e);
2. Every deletable component of Type a) is included in a laminar which is also the first or the last one of a component of Type d) or e);
3. Every iteration of the For loop deleting a vertex deletes or marks as undeletable the vertices of exactly one component of Type e) and all the components of Type a) located in its first and last laminar;
4. Every iteration of the While loop deleting a vertex deletes or marks as undeletable exactly one component of Type d) and all the components of Type a) located in its first and last laminar;

Consequently, FindLaminars terminates with every vertex of $G$ deleted or marked as undeletable.

This result, which proof is postponed to Section 5, is the last one needed to prove the validity of the algorithm.

Lemma 10. The output $(A, Q)$ of FindLaminars is a $(R, K)$ hub-laminar decomposition.

Proof. We shall prove successively all items of the definition of a hub-laminar decomposition (Definition 1).

1. Each laminar links two hubs centers. The endpoints $a, a'$ of any $Q \in Q$ belong to $A$ and for every other hub $a'' \in A \setminus \{a, a'\}$, $B(Q, K) \cap B(a'', R + 1) = \emptyset$: The first part of the claim is straightforward. The second part is a consequence of the last claim of Lemma 4. Indeed, $B(Q, K) \cap B(a'', R + 1) \neq \emptyset$ would imply that the connected component covered by $B(Q, K)$ exhibited three laminars incident to dominated hubs, and thus contained a non-dominated hub of degree at least 3, which is impossible by the first item of Lemma 9.
2. Laminars and hubs dominate $G$: $V(G) = \bigcup_{a \in A} B(a, R) \cup \bigcup_{Q \in Q} B(Q, K)$: This is a direct result of the last claim of Lemma 9.
3. Each laminar path is locally a shortest path. Any path $Q \in Q$ with endpoints $a$ and $a'$ is a shortest path of the graph $G[B(Q, K) \cup B(a, R) \cup$
$B(a', R)$: As it is drawn using a BFS, every path $Q \in \mathcal{Q}$ of endpoints $a$ and $a'$ is a shortest path of the remaining graph when computing $Q$. This graph contains $B(Q, K) \cup B(a, R) \cup B(a', R)$.

4. **Laminars meet at hubs only.** For all $i \neq j$ and $uv \in E(G)$ such that $u \in B(Q_i, K)$ and $v \in B(Q_j, K)$, there is a hub center $a \in A$ such that $Q_i$ and $Q_j$ both have $a$ as endpoint and $u, v \in B(a, R)$.

This is a consequence of the two last items in the enumeration of Lemma 9, which claim that a connected component of deletable vertices cannot stay partially deletable after an iteration of either loop. Indeed, suppose that there exist such vertices $u$ and $v$ that are not in some $B(a, R)$, $a \in A$, and suppose w.l.o.g. that $Q_i$ is added to $\mathcal{Q}$ before $Q_j$. Consider the iteration on which $Q_i$ is built. The connected component of undeletable vertices containing $u$ and $v$ either remains deletable, which contradicts $u \in B(Q_i, K)$, or all non undeletable vertices are deleted, which contradicts $v \notin B(Q_i, K)$.

\[\square\]

The $(K + 2r + k)$-equivalence is a consequence of the fact that $A$ is $H$-dominating, which allows to build the bijection $\phi$ between hub centers with hub degree different from 2. Notice moreover that the 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 5 are met, and the vertex added to $A$ then dominates a hub of $H$. Second, when a vertex of $B$ is transfered to $A$, which happens only when the hub(s) of degree 2 in a problematic configuration have been missed.

Regarding the time complexity, apart from Case (a), each iteration of the while loop in `FindHubs` 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, `FindLaminars` consists in $\lambda$ iterations costing $O(m)$ each.

The code is available on github at [https://github.com/LeoPlanche/hublam]
5. Validity proofs

5.1. Proof of Lemma 3

Proof. The second assertion is straightforward given the first one.

For the sake of contradiction, let us thus assume that there exist a vertex \( w \) on \( P_{u'v'} \) that is not at distance \( 2k \) of \( Q \). For every \( x \in Q_{uv} \), let \( x' \) be a vertex on \( P \) such that \( d(x,x') \leq k \).

As \( u', w \) and \( v' \) are in that order on \( P \), there exist \( x_1 \) which is the closest to \( v \) on \( Q_{uv} \) such that \( x_1', w \) and \( v' \) are in that order. The next vertex \( x_2 \) on \( Q_{uv} \) then verifies that \( x_1', w \) and \( x_2' \) are also on that order on \( P \).

But \( w \) is at distance greater than \( 2k \) of \( x_1 \) and \( x_2 \), so that \( P_{x_1'x_2'} \) is of length at least \( 2k + 2 \). As \( d(x_1', x_2') \leq d(x_1', x_1) + d(x_1, x_2) + d(x_2, x_2') \leq 2k + 1 \), this contradicts the fact that \( P \) is a shortest path.

5.2. Two technical lemmas

Two technical lemmas are needed in order to detail the proofs of the former section’s unproven lemmas. The first one states that a shortest path that enters a laminar but does not traverse it can not enter it deeply, as illustrated in Figure 9.

Lemma 11. Consider a shortest path \( Q \) in the graph induced by \( B(P, k) \) 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 \( m' \) on \( P \), then \( d_G(a, m) \leq 3k \).

Proof.

\[
\begin{align*}
d(m,a) &= d(a,b) - d(m,b) \\
&\leq d(a', b') + 2k - d(m,b) \\
&\leq d(m', b') - d(m', a') + 2k - d(m,b)
\end{align*}
\]

As \( d(m', b') \leq d(m,b) + 2k \) and \( d(m', a') \geq d(m,a) - 2k \), it follows

\[
\begin{align*}
d(m,a) &\leq 6k - d(m,a) \\
&\leq 3k
\end{align*}
\]

\[\square\]
Figure 9: Proof of Lemma 11

The second one states that if some hub $B(h, r)$ is a separator in the connected component covered by StopBFS, the vertex $f$ returned by StopBFS cannot be nearby $h$.

**Lemma 12.** Consider a connected subgraph $g$ of $G$, a hub $B(h, r)$ included in $g$, a vertex $d \in g \setminus B(h, r)$, and a laminar $L$ incident to $B(h, r)$. Let $v_0 = h, v_1, \ldots, v_q$, $q > 2r + 1$, be a subpath of the laminar path of $L$ that belongs to $g$ and suppose that $B(h, r)$ separates $v_q$ from $d$ in $g$. Finally, denote by $f$ the furthest vertex from $d$ in $g$. Then we have $f \notin L \cap B(h, q - 2r - 1)$.

**Proof.** Denote by $S$ the shortest path in $g$ from $d$ to $v_q$. As $B(h, r)$ separates $d$ from $v_q$, $S$ has to hit $B(h, r)$ in some vertex $u$. Then,

$$d_g(d, v_q) = d_g(d, u) + d_g(u, v_q) \geq d_g(d, u) + d_g(h_i, v_q) - d_g(h_i, u) \geq d_g(d, u) + q - r$$

Moreover, for every $w \in B(h_i, q - 2r - 1)$,

$$d_g(d, w) \leq d_g(d, u) + d_g(u, h_i) + d_g(h_i, w) \leq d_g(d, u) + r + q - 2r - 1 < d_g(d, v_q)$$

Thus, $f \notin L \cap B(h, q - 2r - 1)$.

5.3. Proof of Lemma 4

**Proof.** As $r + (K + 2r + k) < R$, we have $B(h, r) \subset B(a, R)$ when $a \in A$ dominates $h \in H$. Conversely, a vertex of $B(a, R)$ cannot hit two different $B(h, r)$ as it would imply $\ell \leq 2(R + r)$. Consequently, as $A$ is $H$-close, $B(h, r)$ does not hit $B(A, R)$ when $h \notin B(A, R)$. Every hub is thus either completely or not at all included in $B(A, R)$.

Consider first a connected component of $G \setminus B(a, R)$ that does not contain a hub. It is therefore included in a laminar. Either it neighbors only one set
$B(a, R)$, corresponding to Type a), or it links some $B(a, R)$ and $B(a', R)$, corresponding to Type d) with $z = 1$.

Consider now the quotient graph where vertices are colored in red if the corresponding hub is included in $B(A, R)$, in black if not. A connected component of $G \setminus B(A, R)$ containing a hub corresponds to a maximal connected subgraph of black vertices. If such a subgraph contains a vertex of degree 3 in the quotient graph, Type b) is met. Otherwise, it corresponds to a path in the quotient graph. Either only one endpoint of that path has a red neighbour, and Type c) is met, or both endpoints have a red neighbour. If the two red vertices are different, this corresponds to Type d) with $z \geq 2$. If the same red vertex neighbors the two endpoints, Type e) is met.

Finally, a laminar $L$ which links two non-dominated hubs is disconnected from the rest of the graph by the two hubs it links, so that no ball $B(a, R)$ can hit or neighbor it, implying the last claim of the lemma.

\[ \square \]

5.4. Proof of Lemma 5

Suppose that there exist a triple $(u, v, w)$ of vertices satisfying the conditions and, for the sake of contradiction, suppose no hub $h$ exists at distance at most $K + 2r + k$ of $u$. $u$ then belongs to some laminar $B(P, k)$, $P \in \mathcal{P}$ linking two hubs $h_1$ and $h_2$ of $H$.

Let us first assume that $w$ does not belong to $B(P, k)$ or belongs to one of the hubs $B(h_1, r)$ or $B(h_2, r)$. The shortest path from $u$ to $v$ then contains a
vertex $x$ of $B(h_1, r)$ or $B(h_2, r)$, so that $d(u, h_1) \leq d(u, x) + d(x, h_1) \leq K + r$ or $d(u, h_2) \leq K + r$. $u$ therefore covers a vertex of $H$.

Assume now that $w \in B(P, k) \setminus (B(h_1, r) \cup B(h_2, r))$. Let $a$ and $b$ be the two vertices such that $Q_{ab} \subset B(P, k)$, $u \in Q_{ab}$ and $Q_{ab}$ is maximal for those two conditions. Then $a$ (resp. $b$) is an endpoint of $Q$ or is a vertex of $B(h_1, r)$ or $B(h_2, r)$. In any case, $d(a, u) \geq K + r + k$ and $d(b, u) \geq K + r + k$.

Let $b'$, $a'$, $w'$ and $u'$ be vertices of $P$ at respective distance at most $k$ from $b$, $a$, $w$ and $u$.

Lemma 11 and the fact that $a$ and $b$ are at distance greater than $3k$ of $u$ imply that $u'$ is between $a'$ and $b'$ on $Q$. Moreover, $w'$ cannot be between $a'$ and $b'$ as Lemma 3 would then imply that $u$ is at distance at most $3k$ of $Q_{a,b}$. We may therefore assume w.l.o.g. that $h_1$, $w'$, $a'$ and $u'$ are on $P$ in that order.

Then $d(u, a) \leq d(u', a') + 2k \leq d(u', w') + 2k \leq d(u, w) + 4k = K + 4k + 1$. As $u \in r_K(Q)$, $a$ is not an endpoint of $Q$, and we may thus assume that $a \in B(h_1, r)$. Then $d(h_1, w') \leq d(h_1, a') - 1 \leq r + k - 1$ and thus $d(u, h_1) \leq d(u, w) + d(w, w') + d(w', h_1) \leq K + 1 + k + r + k - 1 \leq K + 2r + k$. Thus $u$ dominates $h_1$.

Conversely, consider $h \in H$ defining a hub of degree at least 3 and such that StopBFS is rooted outside $B(h, \frac{\ell}{2} - R)$ but explores all this set. Suppose moreover that the path $Q$ output by StopBFS meets the hub $B(h, r)$, $d$ and $f$ beeing outside $B(h, \frac{\ell}{2} - R)$.

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 $r + K + 3k + 2 < \frac{\ell}{2} - R$ from $h$.

Assume first that those three vertices are at distance at most $K$ of vertices $x_i$, $x_j$, $x_l$ in $Q$ respectively. 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 $x_l$ appear in that order in $Q$. There exists therefore a maximal subpath $Q_{ab}$ of $Q$ that is part of $B(P, k) \setminus B(h, r)$ and that contains $x_i$.

Let $a'$ and $b'$ be vertices of $P_i$ such that $d(a, a') \leq k$ and $d(b, b') \leq k$, as illustrated in Figure 11. Then $d(h, a') \leq d(h, a) + k \leq r + k + 1$ and similarly for $b'$. As $d(h, x'_i) > r + k + 1$, Lemma 11 applies to $a$, $x_i$ and $b$ and implies that $d(a, x_i) \leq 3k$ or $d(b, x_i) \leq 3k$. In both cases, as $d(h, a) = d(h, b) = r + 1$ and $d(x_i, x'_i) \leq K$, we get $d(h, x'_i) \leq r + K + 3k + 1$, 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 \( Q \), \( w \) be the following vertex of \( P_i \) and \( u \) be a vertex of \( Q \) such that \( d(u, v) = K \). Then \( d(Q, w) = K + 1 \).

If \( u \) does not belong to the laminar \( B(P_i, k) \), the shortest path from \( u \) to \( v \) has to meet the hub \( B(h, r) \) and \( d(u, v) \leq K + r \). If \( u \in B(P_i, k) \), let \( u' \) be a vertex of \( P_i \) that is at distance at most \( k \) from \( u \). By definition of \( v \), \( u' \) is between \( h \) and \( v \), and \( d(u', v) \geq K - k \) as \( d(u, v) = K \). Thus, \( d(h, u') \leq d(h, v) - K + k \leq d(h, x_i') - K + k \leq r + 4k + 2 \) and \( d(h, u) \leq r + 5k + 2 \). In any case, \( d(h, u) \leq r + K + 2k + 2 \).

Consequently, \( d(u, f) \geq d(h, f) - d(h, u) \geq \frac{\ell}{2} - R - r - K - 2k - 2 > 3K \) and similarly \( d(u, d) \geq 3K \). \( u \) is thus a vertex of \( r_{3K}(Q) \).

5.5. Proof of Lemma 6

Suppose that Invariants 1 to 5 are fulfilled and let \( a \) be a vertex of \( A \). Assume that the set of marked vertices is \( B(a, R + 6k) \setminus B(a, R) \) and that stopBFS is run from a vertex \( d \) such that \( d(a, d) = R + 1 \). Let \( g \) be the
connected component of $G \setminus B(A, R)$ explored by StopBFS, and denote by $Q$ and $f$ the path returned by StopBFS and its last vertex. Note that as $g$ is a subgraph of $G$, $d_g(u, v) \geq d_G(u, v)$ for every vertices $u$ and $v$.

Before dealing with the proof of Lemma 6, let us prove two intermediate results on the regions of $g$ to which $f$ cannot belong. To do so, let $H_0$ denote the hub whose center $h_0 \in H$ is dominated by $a$, $L_1$ the laminar containing $d$ which is incident to $H_0$, and $H_1$ the other hub $L_1$ is incident to, while $h_1$ denotes its center.

The first intermediate lemma concerns the behavior of StopBFS in $L_1$, as the distances in $g$ and $G$ may be quite different there.

**Lemma 13.** Suppose that StopBFS explores an uncolored vertex. Then one of the following claims hold:

1. $g$ is of Type c) with $z = 1$ and $f$ has a colored neighbor;
2. $f$ dominates $h_1$;
3. $f$ is outside $L_1 \cup H_1$;

In particular, the explored component is not of Type a).

**Proof.** Suppose that none of the claims is true. $h_1$ is then not dominated, otherwise the first claim would hold. Moreover, as $H_1 \subset B(h_1, K + 2r + k)$, $f$ is a vertex of $L_1 \setminus H_1$.

Let $u$ be the first uncolored vertex of $Q$, and let $d', u'$ and $f'$ be vertices of $P_1$ that are at respective distances less than $k$ from $d$, $u$ and $f$.

Suppose first that $f \in B(a, R + K)$, as illustrated in Figure 12. Then, as $d_G(a, d) = R + 1$ and $d_G(a, u) \geq R + 2K + 2$, $Q$ contains two vertices $v_1$ and $v_2$ such that $d_G(a, v_1) = d_G(a, v_2) = R + K + 1$ and $u \in Q_{v_1 v_2}$. Moreover, $d_g(v_i, u) \geq d_G(v_i, u) - d_G(a, v_i) = K + 1, 1 \leq i \leq 2$. The vertices $v_1'$ and $v_2'$ on $P_1$ that are at distance at most $k$ of $v_1$ and $v_2$ are then both closer to $h_0$ than $u'$. As $v_1$ and $v_2$ are at distance at least $K + 1$ of $B(a, R)$, $Q_{v_1 v_2}$ is also a shortest path between them in $G$, so that this configuration contradicts Lemma 11. Thus $f$ is not in $B(a, R + K)$.

The former paragraph implies that $d'$, $f'$ and $h_1$ are in that order on $P_1$. Lemma 3 thus implies that the shortest path in $g$ linking $d$ to $h_1$ $3k$ covers $f$: there exist $x$ on that path such that $d_G(x, f) \leq 3k$. Note first that, as $f \notin B(a, R + K)$, none of the vertices on the shortest path from $f$ to $x$ belongs to $B(a, R)$, and the same holds for the subpath of $P_1$ from $x$ to $h_1$. Thus, $h_1$ belongs to the same component than $f$, that is $g$. This implies that $g$ is not of Type a).
Furthermore, as $f$ is at distance more than $3k$ of $B(a, R)$, $d_G(x, f) \leq 3k$ implies $d_g(x, f) \leq 3k$. Thus, as $f$ is the furthest vertex from $d$ in $g$ and $x$ in on the shortest path from $d$ to $h_1$, $d_g(x, h_1) \leq 3k$. Finally, $d_G(h_1, f) \leq d_g(h_1, f) \leq 6k \leq K + 2r + k$ and $f$ dominates $h_1$.

Figure 13 illustrates the notations used in the previous paragraph.

![Figure 13: Notations used in the first part of Lemma 13. The shaded region corresponds to marked vertices.](image)

**Lemma 14.** Suppose that StopBFS explores an uncolored vertex. Consider a maximal sequence of laminars and hubs $L_1, H_1, \ldots, L_z, H_z, L_{z+1}$, such that all hubs are in $g$ and every hub $H_i$ with $i \leq z - 1$ is of degree 2. Denote by $v$ the last vertex of the laminar path $P_{z+1}$ of $L_{z+1}$ that belongs to $g$ (if $L_{z+1}$ is entirely included in $g$, $v$ is the hub-center of the second hub $L_{z+1}$ is incident to).

Then either $f \in B(v, 6k)$, or $f$ does not belong to $\bigcup_{1 \leq i \leq z} H_i \cup \bigcup_{1 \leq i \leq z+1} L_i$.

**Proof.** If $z = 0$, the sequence is limited to $L_1$ and Lemma 13 applies. Consider $z \geq 1$. By Lemma 13, $f$ doesn’t belong to $L_1$.

Consider a hub $H_i$ of center $h_i$, $1 \leq i \leq z$ and let $x$ be a vertex on the middle of the laminar path $P_{i+1}$ of $L_{i+1}$. The part of $P_{i+1}$ between $h_i$ and $x$ belongs to $g$ and, as all hubs from $H_1$ to $H_{i-1}$ are of degree 2, $H_i$ separates $d$ from $x$. By Lemma 12 and as $r + 6k < \ell - 2r - 1$, $f$ does therefore not belong to $\bigcup_{1 \leq i \leq z} B(h_i, r + 6k)$.  

32
Suppose now that $z \geq 2$ and consider a vertex $u$ which is in $L_i \setminus (B(h_{i-1}, r + 6k) \cup B(h_i, r + 6k))$. Let $S$ be the shortest path in $g$ from $d$ to $h_i$, and $w$ a vertex in $S \cap H_{i-1}$. Such a vertex has to exist as there are only hubs of degree 2 between $L_1$ and $L_i$. Let $u'$ and $w'$ be vertices of $P_i$ at distance less than $k$ respectively from $u$ and $w$. Then

$$d(h_{i-1}, w') \leq d(h_{i-1}, w) + d(w, w') \leq r + k$$

and

$$d(h_{i-1}, u') \geq d(h_{i-1}, u) - d(u, u') \geq r + 5k$$

Thus, $u'$ is between $w'$ and $h_i$ on $P_i$, so that by Lemma 3 there exist $x$ on $S$ such that $d_g(x, u) = d_G(x, u) \leq 3k$. Then

$$d_g(d, h_i) = d_g(d, x) + d_g(x, h_i)$$
$$\geq d_g(d, x) + d_g(x, u) - 3k + d_g(x, h_i)$$
$$\geq d_g(d, u) - 3k + d_g(x, h_i)$$

Moreover, as $x$ is at distance $3k$ of $u$ and $u$ at distance at least $6k + 1$ of $h_i$, $d_g(x, h_i) > 3k$. Finally, $d_g(d, h) > d_g(d, u)$, implying that $u$ cannot be the vertex at greatest distance from $d$. $f$ is therefore not a vertex of $L_i$.

Figure 14 illustrates the notations used in the previous paragraph.
Finally, let \( u \) be a vertex in \( L_{z+1} \setminus (B(h_z, r + 6k) \cup B(v, 6k)) \). The former paragraph can be mimiced by replacing \( h_i \) by \( v \): the shortest path \( S \) in \( g \) from \( d \) to \( v \) has to 3\( k \)-cover \( u \), which implies that \( u \) is closer than \( v \) to \( d \) because \( d_g(u, v) \geq 6k + 1 \). \( f \) is therefore not in \( L_{z+1} \setminus B(v, 6k) \).

Let us now prove Lemma \( \ref{lemma:6} \) by considering the different topologies listed in Lemma \( \ref{lemma:4} \).

**Type a)** Lemma \( \ref{lemma:13} \) implies that StopBFS explores no uncolored vertices.

**Type b)** Following the sequence of incident laminars and hubs starting at \( L_1 \) until a hub of degree at least 3 is met, \( g \) contains a sequence corresponding to Lemma \( \ref{lemma:14} \) with \( H_z \) of degree at least 3.

Either \( h_{z+1} \) is not dominated, and \( f \) is either outside \( L_{z+1} \) or dominates \( h_{z+1} \) by Lemma \( \ref{lemma:14} \). Or \( h_{z+1} \) is dominated and the last uncolored vertex \( v \) on \( P_{z+1} \) is both at distance at most 6\( k \) of \( f \), again by Lemma \( \ref{lemma:14} \) and at distance and most \( R + (K + 2r + k) \) of \( h_{z+1} \). In any case, \( f \) is at distance at least \( \ell - R - K - 2r - 7k > \frac{\ell}{2} - R \) from \( h \).

Consequently, Lemma \( \ref{lemma:5} \) implies that a triple \((u, v, w)\) satisfying its conditions has to exist.

**Type c)** If \( z = 1 \), Lemma \( \ref{lemma:13} \) ensures that \( f \) dominates \( h_1 \). Moreover, \( d_G(h_0, h_1) \geq \ell \) then ensures that \( f \) is not marked.

If \( z \geq 2 \), Lemma \( \ref{lemma:14} \) implies that \( f \) is in \( B(h_z, 6k) \). As \( 6k \leq K + 2r + k \), \( f \) dominates \( h_z \). \( d_G(h_0, h_z) \geq \ell \) then implies that \( f \) is not marked.

The definition of the Type c) component moreover implies that \( f \) cannot neighbor a colored vertex.
Type d) Let \( v \) be the last vertex on \( P_z \) which belongs to \( g \). As \( L_z \) is incident to \( B(h_z, r) \), \( h_z \neq h_0 \), the next vertex on \( P_z \) is not in \( g \) because it belongs to \( B(a', R) \), \( a' \neq a \). When exploring \( g \), StopBFS therefore explores a vertex which neighbors \( B(a', R) \) so that it stops with \( f \) being such a vertex. \( f \) thus has a colored neighbor.

Moreover, \( d(a, f) \geq d(a, a') - d(a', f) \geq d(h_0, h') - d(h_0, a) - d(h', a') - d(a'f) \geq \ell - 2(K + 2r + k) > 6k \). \( f \) is thus uncolored.

Type e) Let \( v \) be the last vertex on \( P_z \) which belongs to \( g \). As \( L_z \) is incident to \( B(h_0, r) \), the next vertex on \( P_z \) is not in \( g \) because it belongs to \( B(a, R) \). Lemma 14 moreover implies that \( f \) belongs to \( B(v, 6k) \).

Finally, \( d_G(a, f) \leq R + 1 + 6k \), so that \( f \) is a marked vertex.

Moreover, the path from \( d \) to \( f \) on the StopBFS tree has to cross \( B(h_1, r) \), so that it contains an unmarked vertex. \( f \) is thus different from \( d \).

5.6. Proof of Lemma 8

Let \( Q \) be the path computed by the BFS from \( s \) to \( x \). Assume that \( Q \) intersects a hub of degree at least 3 and of center \( h \). Let \( y \) be a vertex in this intersection. Furthermore assume that \( y \) is at distance more than \( 4K + 3k + 2 + 2r \) of \( s \) and \( x \). Those vertices are then at distance more than \( 4K + 3k + 2 + r \) of \( h \), hence not in \( B(h, 4K + 3k + 2 + r) \). No vertex of \( G \) being colored, the BFS starting on \( s \) contains \( G \) and a fortiori \( B(h, 4K + 3k + 2 + r) \). Every condition of lemma 5 is verified, so that a triplet of vertices \((u, v, w)\) is detected.

Assume now that no triplet as defined in lemma 5 is found. By the last paragraph we know that no vertex at distance more than \( 4K + 3k + 2 + 2r \) of \( s \) and \( x \) is in a hub of degree 3 or more. The path \( Q_{s+4K+3k+2+2r,x-4K+3k+2+2r} \) is therefore in an unique laminar or in an alternating sequence of laminars and hubs of degree 2.

If \( s \) is in a laminar, consider \( h \) an hub center which is not an extremity of that laminar. If \( s \) is in a hub, consider \( h \) an hub center different from the one of \( s \). In any case, a path from \( s \) to \( h \) starts or goes through a hub \( H' \) with \( h' \neq h \), we have :

\[
d(s, h) \geq d(h, h') - 2r \geq \ell - 2r
\]
The path $Q$ is then of size at least $\ell - 2r$. Let $m$ by in the middle of $Q$. By the preceding remarks $m$ is at distance at least $\frac{\ell}{2} - 4K - 3k - 2r - 2$ of every hub of degree 3 or more. The vertex $m$ is in the center of a sequence $S = H_0, L_0, ... H_i, L_i, ... H_z$ such that every hub of $S$ is of degree 2, expect the extremies $H_0$ and $H_z$. $G$ being connex and containing an hub of degree at least 3, at least one extremity of the sequence is of degree 3 or more. Note that we may have $H_0 = H_z$.

If $m$ is in an hub $H$, then $B(m, R)$ contains $H$ and disconnects $S$. Let $d$ by a vertex of $Q$ at distance $R + 1$ of $m$ the closest from $s$. We have $d$ at distance at least $\frac{\ell}{2} - 4K - 3k - R - 2r - 3$ of an hub of degree 3 or more. Furthermore $d$ is in a sequence $S' = H_0, L_0, ... H_i, L_i, ... B(m, R)$ such that $H_0$ is of degree 1 or of degree 2 or more and disconnects $g = G \setminus B(m, R)$.

In this second case, by Lemma 12, the BFS in $g$ starting on $d$ and reaching $f$ detects a triplet of vertices $(u, v, w)$ as defined by Lemma 5.

We now only have to show that $H_0$ is not an hub of degree 1. Assume the opposite. The vertex $s$ is then in a sequence $S = H_0, L_0, ... H_i, L_i, ... H_z$ with $H_z$ of degree at least 3, disconnecting $S$ from the rest of the graph. If $x$ is in $S$ then $B(x, R)$ disconnects $s$ from $G \setminus S$. Let $h$ be an hub outside of $S$,

$$d(s, h) \geq d(s, x) - 2R + d(x, h) \geq d(s, x) - 2R + d(h_z, h) - 2r$$

$$\geq d(s, x) + \ell - 2(R + r) > d(s, x)$$

It contradicts the fact that $x$ is a vertex furthest from $s$. If $x$ is not in $S$, it is still at distance at most $4K + 3k + 2 + 2r$ from $h_z$.

$$d(s, h) \geq d(s, h_z) + d(h_z, h') - 2r \geq d(s, h_z) + \ell - 2r$$

$$d(s, x) \leq d(s, h_z) + d(h_z, x) + 2r \leq d(s, h_z) + 4K + 3k + 2 + 4r \leq d(s, h_z) + \ell - 2r$$

A contradiction is again obtained to the the fact that $x$ is a vertex furthest from $s$.

5.7. Proof of Lemma 9

The two first items are verified at the beginning of the algorithm. Indeed, the set $A$ returned by FindHubs is $H$-dominating. Therefore, all hubs of degree 1 or 3 are dominated and thus only components of Type a), d) or e) are
present in \( G \setminus B(A,R) \). Moreover, every component of Type a) is contained in some \( B(a,R+2K+1) \), \( a \in A \), as a consequence of Lemma \[13\]. As \( \ell > R \), the central vertex \( x \) of the laminar path containing a Type a) component is deletable. Moreover, \( d(a,x) = d(h,x) - d(a,h) \geq \lfloor \ell/2 \rfloor - (K + 2r + k) > R + 2K + 1 \), so that \( x \) has to belong to a Type d) or e) component.

Once initially true, those two items clearly remain true given the two last ones for each iteration of either loop.

Moreover, at each iteration the number of deletable vertices decreases strictly until there are no more components of Type d) or e), the algorithm ends with no deletable vertices. It is therefore sufficient to show that the two last items are verified given the first ones.

Consider an iteration of the For loop that deletes vertices. As \( \text{FindHubs} \) added vertices in \( B \) only in the middle of Type e) components, \( b \) belongs to such a component.

All vertices of that component, as well as all vertices of Type a) components included in \( L_1 \) or \( L_z \), are \( K \)-covered by \( Q_1 \) or \( Q_2 \). Indeed, if it were not the case, an uncolored vertex \( c \) is found and added to \( A \) at Line \[15\] and no vertex is deleted in that loop.

To prove that no other vertices from other components are deleted, we have to prove that \( Q_1 \) (and by symmetry \( Q_2 \)) does not \( K \)-cover any vertex of a laminar \( L \) incident to \( B(a,R) \) and different from \( L_1 \). Let \( h \) be the vertex of \( H \) dominated by \( a \), and such \( L \) and \( L_1 \) are incident to \( B(h,r) \).

The deletable vertices of \( L \) being at distance at least \( R - (K + 2r + k) \geq r + 3k + K \) from \( h \), none of them is deleted if \( Q \) does not enter \( L \).

Let therefore assume there exits vertices \( x, y \) and \( z \) appearing on \( Q \) in this order, such that \( x \) and \( z \) are \( L \cap B(h,r) \) and \( y \) is the vertex on \( Q \) the furthest of \( h \). Let \( x', y' \) and \( z' \) be vertices on the laminar path \( k \)-covering them. If \( y' \) is closer to \( h \) than \( x' \) or \( z' \), say \( x' \),

\[
d(h_1,y) \leq d(h_1,x') + d(y',y) \leq d(h_1,x) + d(x,x') + d(y',y) \leq r + 2k
\]

If not, Lemma \[11\] implies that \( d(x,y) \leq 3k \). In any case, \( d(h,y) \leq r + 3k \). Thus any vertex \( K \)-covered by \( Q \) is at distance at most \( r + 3k + K \) from \( h \), that is is undeletable. None of the deletable vertices of \( L(x) \) is thus deleted.

Consider now an iteration of the While loop that deletes vertices. As all Type e) components contained some vertex \( b \in B \) and where therefore deleted during the for loop, only components of Type a) and d) remain.
Every constructed path $Q$ linking some $B(a, R)$ to $B(a', R)$, $a' \leq a$, it hits a component of Type d). All vertices of that component, as well as all vertices of Type a) components included in $L_1$ or $L_2$, are then $K$-covered by $Q$. Indeed, suppose it is not the case, and let $w$ be a vertex of such a component at distance $K+1$ of $Q$. Consider $u$ on $Q$ such that $d(u, w) = K+1$. As $R \geq 4K + 2$ and $w \notin B(a, R)$, $u$ belongs to $r_{3K}(Q)$. A triple satisfying Lemma 5 is thus found and no vertex is deleted.

The fact that no vertices in other components are deleted in that iteration is proven in the same way as for the For loop.

6. Embedding and distance labeling

6.1. Circle embedding with bounded distortion

Proposition 1, stated in Section 3, is a consequence of Theorem 1 and the two following propositions.

Proposition 3. Any graph $G$ having a circle embedding with distortion $\gamma$ has a shortest path or an isometric cycle with eccentricity $\lfloor \gamma/2 \rfloor$ at most.

Proof. Consider an embedding of $G$ in a circle $C$ with distortion $\gamma$. Suppose that any shortest path of $G$ has eccentricity greater than $\lfloor \gamma/2 \rfloor$. We first show that $G$ contains a simple cycle that $\lfloor \gamma/2 \rfloor$-dominates the graph. Given a path $P$, two consecutive nodes $u, v$ of $P$ are at distance at most $\gamma$ in the circle embedding, and $P$ thus $\lfloor \gamma/2 \rfloor$-dominates any node embedded between $u$ and $v$ in the circle. We define the arc $P_C$ of $P$ in $C$ as the smallest arc of $C$ where nodes of $P$ are embedded. Note that all nodes embedded in $P_C$ are $\lfloor \gamma/2 \rfloor$-dominated by $P$. Consider a shortest path $P$ with longest arc $P_C$ and let $a, b$ denote the extremities of $P_C$. If $P$ does not $\lfloor \gamma/2 \rfloor$-dominate $G$, consider a node $c$ at distance greater than $\lfloor \gamma/2 \rfloor$ from $P$. Node $c$ cannot be embedded in $P_C$. Consider a shortest path $Q$ from $c$ to $a$ in $G$. The arc $Q_C$ contains one of the two circle arcs joining $c$ and $a$. The choice of $P$ implies that $Q_C$ cannot contain $P_C$. The path $Q_C$ thus $\lfloor \gamma/2 \rfloor$-dominates nodes embedded in the arc $C_{ca}$ of $C$ from $c$ to $a$ that avoids the interior of $P_C$. Similarly, the shortest path $R$ from $c$ to $b$ dominates nodes embedded in the arc $C_{cb}$ of $C$ from $c$ to $b$ that avoids the interior of $P_C$. Let $a'$ be the first node of $Q$ in $P$. Let $Q'$ be the sub-path of $Q$ from $c$ to $a'$ and let $P'$ be the sub-path of $P$ from $a'$ to $b$. Note that the arc of $Q' \cup P'$ contains the arc in $C$ from $c$ to $b$ in $Q_C \cup P_C$. Similarly, let $b'$ be the first node of $R$ in $Q' \cup P'$.
Then define $R'$ as the sub-path of $R$ from $c$ to $b'$ and $Q''$ as the sub-path of $Q' \cup P'$ from $c$ to $b'$. Note that $R'_C$ contains the arc from $c$ to $b$ which is not in $R_C \cup P_C$. The union $Q'' \cup R'$ defines a simple cycle that $\lfloor \gamma/2 \rfloor$-dominates $G$ as $Q'_C \cup R'_C = C$.

Now consider a simple cycle $S$ of $G$ that $\lfloor \gamma/2 \rfloor$-dominates $G$ and has minimum length. $S$ must be isometric: otherwise there would be a path $P$ from $a$ to $b$ in $G$ that is shorter than both paths $Q$ and $R$ of $S$ from $a$ to $b$. Consider the arc $A$ of $C$ from $a$ to $b$ included in $P_C$. Without loss of generality, $Q$ dominates the nodes embedded in the other part $C \setminus A$ of the cycle. We can then construct from $P \cup Q$ (similarly as above) a simple cycle that $\lfloor \gamma/2 \rfloor$-dominates $G$ in contradiction with the choice of $S$ as $|P| + |Q| < |S|$. 

**Proposition 4.** Given a graph $G$ and an isometric cycle with eccentricity $k$ in $G$, an embedding of $G$ in a circle with distortion $O(k \cdot cd(G))$ can be computed in polynomial time.

**Proof.** The construction of the embedding is similar to that of Dragan and Leitert (2015) with Euler tours of trees of depth $k$ rooted. However, our trees are rooted on a cycle rather than a path. Consider an isometric cycle $C$ of $G$ having eccentricity $k$. We construct a forest $F$ with roots in $C$ as a union of shortest paths: for each node $u \in V(G)$ we select a node $u'$ such that $d(u, u') = d(u, C)$ and add to $F$ a shortest path from $u$ to $u'$ ($u' = u$ for $u \in C$). For each tree $T$ of $F$ rooted at a node $c \in C$, we construct an Euler tour $E_c$ which is a sequence of tree edges starting from $c$, visiting all nodes of $T$ in a depth-first-search manner and terminating at $c$. Each edge is used twice and the length of $E_c$ is $2(n' - 1)$ where $n'$ denote the number of nodes in $T$. We then obtain a tour of the whole graph as the sequence $E_C = E_{c_1}, c_1 c_2, E_{c_2}, \ldots, c_{p-1} c_p, E_{c_p}, c_p c_1$ where $p$ is the length of $C$ and $c_1, \ldots, c_p$ are the nodes of $C$ ordered according to the cycle order. Note that this tour contains $2n$ edges at most and can be embedded in a circle $C'$ with same length.

We now analyze the distortion of this circle embedding in $C'$. Given an edge $uv$ of $G$, consider the roots $u'$ and $v'$ of the trees of $u$ and $v$ respectively. Let $S$ denote the union of trees rooted on the shortest path from $u'$ to $v'$ in $C$. Note that the distance from $u$ to $v$ in the tour $E_C$ is at most twice the size of $S$. To upper-bound $|S|$, we consider an embedding of $G$ in a circle $C_{opt}$ with distortion $\gamma = cd(G)$. As we have $d(u', v') \leq 2k + 1$, the diameter of $S$ is at most $4k+1$. Two nodes of $S$ are thus embedded at distance at most $\gamma(4k+1)$
in the circle \( C_{opt} \) and different nodes are at distance 1 at least. We thus have \( |S| \leq 2\gamma(4k + 1) \), and our embedding in \( C' \) has distortion \( O(\gamma k) \). \( \square \)

6.2. Distance labeling for general hub-laminar decomposition

A hub-laminar 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 \( \hat{d}_{uv} \) can be computed from \( L_u \) and \( L_v \) such that \( d(u, v) \leq \hat{d}_{uv} \leq d(u, v) + c \). Proposition 2 is a consequence of Theorem 2 and the following proposition.

**Proposition 5.** Given a \((r, k)\)-hub-laminar decomposition with \( \lambda \) laminars \((H, P)\) of a graph \( G \), a \( \max(4k, 2r) \)-additive distance labeling with \( O(\lambda \log n) \) bit labels can be computed in polynomial time.

**Proof.** We assume that hub centers are numbered from 1 to \( q \), \( q \leq 2\lambda \) and laminars from 1 to \( \lambda \). 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 with number \( \alpha \), i.e. there exists \( P \in \mathcal{P} \) with endpoints \( h_1 < h_2 \) such that \( u \in B(P, k) \), we additionally store \((d_P(h_1, u'), d(u', u), \alpha)\) for some \( u' \in B(u, k) \cap P \) and set \( L_u := (d_P(h_1, u'), d(u', u), \alpha), H_u \) (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 triples \((d(h_1, u'), d(u', u), \alpha)\) and \((d(h_1, v'), d(v', v), \alpha)\) respectively with \( \alpha = \alpha' \), 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 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 Axioms 2 and 4 of Definition 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 triples \((d(h_1, u'), d(u', u), \alpha)\) and \((d(h_1', v'), d(v', v), \alpha')\) respectively. Consider the subgraph \(G_P\) induced by \(B(P, k)\). By triangle inequality, 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 \(G_P\) 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 \(f(u, v) \leq d(u, v) + \max(4k, 2r)\). \(\square\)

7. 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 and enlightening the possible biological meanings of hubs.

References


URL http://doi.acm.org/10.1145/1060590.1060624

URL http://dl.acm.org/citation.cfm?id=1070432.1070449


URL http://doi.acm.org/10.1145/1044731.1044732

URL https://hal.inria.fr/hal-01282715