

# Steiner trees and higher geodecity

Daniel Weißauer

## Abstract

Let  $G$  be a connected graph and  $\ell : E(G) \rightarrow \mathbb{R}^+$  a length-function on the edges of  $G$ . The *Steiner distance*  $\text{sd}_G(A)$  of  $A \subseteq V(G)$  within  $G$  is the minimum length of a connected subgraph of  $G$  containing  $A$ , where the length of a subgraph is the sum of the lengths of its edges.

It is clear that every subgraph  $H \subseteq G$ , with the induced length-function  $\ell|_{E(H)}$ , satisfies  $\text{sd}_H(A) \geq \text{sd}_G(A)$  for every  $A \subseteq V(H)$ . We call  $H \subseteq G$  *k-geodesic in G* if equality is attained for every  $A \subseteq V(H)$  with  $|A| \leq k$ . A subgraph is *fully geodesic* if it is *k-geodesic* for every  $k \in \mathbb{N}$ . It is easy to construct examples of graphs  $H \subseteq G$  such that  $H$  is *k-geodesic*, but not  $(k+1)$ -geodesic, so this defines a strict hierarchy of properties. We are interested in situations in which this hierarchy collapses in the sense that if  $H \subseteq G$  is *k-geodesic*, then  $H$  is already fully geodesic in  $G$ .

Our first result of this kind asserts that if  $T$  is a tree and  $T \subseteq G$  is 2-geodesic with respect to some length-function  $\ell$ , then it is fully geodesic. This fails for graphs containing a cycle. We then prove that if  $C$  is a cycle and  $C \subseteq G$  is 6-geodesic, then  $C$  is fully geodesic. We present an example showing that the number 6 is indeed optimal.

We then develop a structural approach towards a more general theory and present several open questions concerning the big picture underlying this phenomenon.

## 1 Introduction

Let  $G$  be a graph and  $\ell : E(G) \rightarrow \mathbb{R}^+$  a function that assigns to every edge  $e \in E(G)$  a positive *length*  $\ell(e)$ . This naturally extends to subgraphs  $H \subseteq G$  as  $\ell(H) := \sum_{e \in E(H)} \ell(e)$ . The *Steiner distance*  $\text{sd}_G(A)$  of a set  $A \subseteq V(G)$  is defined as the minimum length of a connected subgraph of  $G$  containing  $A$ , where  $\text{sd}_G(A) := \infty$  if no such subgraph exists. Every such minimizer is necessarily a tree and we say it is a *Steiner tree for A in G*. In the case where  $A = \{x, y\}$ , the Steiner distance of  $A$  is the ordinary distance  $d_G(x, y)$  between  $x$  and  $y$ . Hence this definition yields a natural extension of the notion of “distance” for sets of more than two vertices. Corresponding notions of radius, diameter and convexity have been studied in the literature [6, 2, 1, 3, 9, 5]. Here, we initiate the study of *Steiner geodecity*, with a focus on structural assumptions that cause a collapse in the naturally arising hierarchy.

Let  $H \subseteq G$  be a subgraph of  $G$ , equipped with the length-function  $\ell|_{E(H)}$ . It is clear that for every  $A \subseteq V(H)$  we have  $\text{sd}_H(A) \geq \text{sd}_G(A)$ . For a natural number  $k$ , we say that  $H$  is *k-geodesic in G* if  $\text{sd}_H(A) = \text{sd}_G(A)$  for every  $A \subseteq V(H)$  with  $|A| \leq k$ . We call  $H$  *fully geodesic in G* if it is *k-geodesic* for every  $k \in \mathbb{N}$ .

By definition, a *k-geodesic* subgraph is *m-geodesic* for every  $m \leq k$ . In general, this hierarchy is strict: In Section 6 we provide, for every  $k \in \mathbb{N}$ , examples of graphs  $H \subseteq G$  and a length-function  $\ell : E(G) \rightarrow \mathbb{R}^+$  such that  $H$  is *k-geodesic*, but not *(k + 1)-geodesic*. On the other hand, it is easy to see that if  $H \subseteq G$  is a 2-geodesic *path*, then it is necessarily fully geodesic, because the Steiner distance of any  $A \subseteq V(H)$  in  $H$  is equal to the maximum distance between two  $a, b \in A$ . Our first result extends this to all trees.

**Theorem 1.1.** *Let  $G$  be a graph with length-function  $\ell$  and  $T \subseteq G$  a tree. If  $T$  is 2-geodesic in  $G$ , then it is fully geodesic.*

Here, it really is necessary for the subgraph to be acyclic (see Corollary 6.5). Hence the natural follow-up question is what happens in the case where the subgraph is a cycle.

**Theorem 1.2.** *Let  $G$  be a graph with length-function  $\ell$  and  $C \subseteq G$  a cycle. If  $C$  is 6-geodesic in  $G$ , then it is fully geodesic.*

Note that the number 6 cannot be replaced by any smaller integer.

In Section 2 we introduce notation and terminology needed in the rest of the paper. Section 3 contains observations and lemmas that will be used later. We then prove Theorem 1.1 in Section 4. In Section 5 we prove Theorem 1.2 and provide an example showing that the number 6 is optimal. Section 6 contains an approach towards a general theory, aiming at a deeper understanding of the phenomenon displayed in Theorem 1.1 and Theorem 1.2. Finally, we take the opportunity to present the short and easy proof that in any graph  $G$  with length-function  $\ell$ , the cycle space of  $G$  is generated by the set of fully geodesic cycles.

## 2 Preliminaries

All graphs considered here are finite and undirected. It is convenient for us to allow parallel edges. In particular, a cycle may consist of just two vertices joined by two parallel edges. Loops are redundant for our purposes and we exclude them to avoid trivialities. Most of our notation and terminology follows that of [7], unless stated otherwise.

A set  $A$  of vertices in a graph  $G$  is called *connected* if and only if  $G[A]$  is.

Let  $G, H$  be two graphs. A *model of G in H* is a family of disjoint connected *branch-sets*  $B_v \subseteq V(H)$ ,  $v \in V(G)$ , together with an injective map  $\beta : E(G) \rightarrow E(H)$ , where we require that for any  $e \in E(G)$  with endpoints  $u, v \in V(G)$ , the edge  $\beta(e) \in E(H)$  joins vertices from  $B_u$  and  $B_v$ . We say that  $G$  is a *minor* of  $H$  if  $H$  contains a model of  $G$ .

We use additive notation for adding or deleting vertices and edges. Specifically, let  $G$  be a graph,  $H$  a subgraph of  $G$ ,  $v \in V(G)$  and  $e = xy \in E(G)$ . Then  $H + v$  is the graph with vertex-set  $V(H) \cup \{v\}$  and edge-set  $E(H) \cup \{vw \in E(G) : w \in V(H)\}$ . Similarly,  $H + e$  is the graph with vertex-set  $V(H) \cup \{x, y\}$  and edge-set  $E(H) \cup \{e\}$ .

Let  $G$  be a graph with length-function  $\ell$ . A *walk* in  $G$  is an alternating sequence  $W = v_1 e_1 v_2 \dots e_k v_{k+1}$  of vertices  $v_i$  and edges  $e_i$  such that  $e_i = v_i v_{i+1}$  for every  $1 \leq i \leq k$ . The walk  $W$  is *closed* if  $v_1 = v_{k+1}$ . Stretching our terminology slightly, we define the *length* of the walk as  $\text{len}_G(W) := \sum_{1 \leq i \leq k} \ell(e_i)$ . The *multiplicity*  $m_W(e)$  of an edge  $e \in E(G)$  is the number of times it is traversed by  $W$ , that is, the number of indices  $1 \leq j \leq k$  with  $e = e_j$ . It is clear that

$$\text{len}_G(W) = \sum_{e \in E(G)} m_W(e) \ell(e). \quad (1)$$

Let  $G$  be a graph and  $C$  a cycle with  $V(C) \subseteq V(G)$ . We say that a walk  $W$  in  $G$  is *traced by*  $C$  in  $G$  if it can be obtained from  $C$  by choosing a starting vertex  $x \in V(C)$  and an orientation  $\vec{C}$  of  $C$  and replacing every  $\vec{ab} \in E(\vec{C})$  by a shortest path from  $a$  to  $b$  in  $G$ . A cycle may trace several walks, but they all have the same length: Every walk  $W$  traced by  $C$  satisfies

$$\text{len}_G(W) = \sum_{ab \in E(C)} d_G(a, b). \quad (2)$$

Even more can be said if the graph  $G$  is a tree. Then all the shortest  $a$ - $b$ -paths for  $ab \in E(C)$  are unique and all walks traced by  $C$  differ only in their starting vertex and/or orientation. In particular, every walk  $W$  traced by  $C$  in a tree  $T$  satisfies

$$\forall e \in E(T) : m_W(e) = |\{ab \in E(C) : e \in aTb\}|, \quad (3)$$

where  $aTb$  denotes the unique  $a$ - $b$ -path in  $T$ .

Let  $T$  be a tree and  $X \subseteq V(T)$ . Let  $e \in E(T)$  and let  $T_1^e, T_2^e$  be the two components of  $T - e$ . In this manner,  $e$  induces a bipartition  $X = X_1^e \cup X_2^e$  of  $X$ , given by  $X_i^e = V(T_i^e) \cap X$  for  $i \in \{1, 2\}$ . We say that the bipartition is *non-trivial* if neither of  $X_1^e, X_2^e$  is empty. The set of leaves of  $T$  is denoted by  $L(T)$ . If  $L(T) \subseteq X$ , then every bipartition of  $X$  induced by an edge of  $T$  is non-trivial.

Let  $G$  be a graph with length-function  $\ell$ ,  $A \subseteq V(G)$  and  $T$  a Steiner tree for  $A$  in  $G$ . Since  $\ell(e) > 0$  for every  $e \in E(G)$ , every leaf  $x$  of  $T$  must lie in  $A$ , for otherwise  $T - x$  would be a tree of smaller length containing  $A$ .

In general, Steiner trees need not be unique. If  $G$  is a tree, however, then every  $A \subseteq V(G)$  has a unique Steiner tree given by  $\bigcup_{a, b \in A} aTb$ .

### 3 The toolbox

The first step in all our proofs is a simple lemma that guarantees the existence of a particularly well-behaved substructure that witnesses the failure of a subgraph to be  $k$ -geodesic.

Let  $H$  be a graph,  $T$  a tree and  $\ell$  a length-function on  $T \cup H$ . We call  $T$  a *shortcut tree for  $H$*  if the following hold:

$$(SCT1) \quad V(T) \cap V(H) = L(T),$$

$$(SCT2) \quad E(T) \cap E(H) = \emptyset,$$

$$(SCT3) \quad \ell(T) < \text{sd}_H(L(T)),$$

$$(SCT4) \quad \text{For every } B \subsetneq L(T) \text{ we have } \text{sd}_H(B) \leq \text{sd}_T(B).$$

Note that, by definition,  $H$  is not  $|L(T)|$ -geodesic in  $T \cup H$ .

**Lemma 3.1.** *Let  $G$  be a graph with length-function  $\ell$ ,  $k$  a natural number and  $H \subseteq G$ . If  $H$  is not  $k$ -geodesic in  $G$ , then  $G$  contains a shortcut tree for  $H$  with at most  $k$  leaves.*

*Proof.* Among all  $A \subseteq V(H)$  with  $|A| \leq k$  and  $\text{sd}_G(A) < \text{sd}_H(A)$ , choose  $A$  such that  $\text{sd}_G(A)$  is minimum. Let  $T \subseteq G$  be a Steiner tree for  $A$  in  $G$ . We claim that  $T$  is a shortcut tree for  $H$ .

*Claim 1:*  $L(T) = A = V(T) \cap V(H)$ .

The inclusions  $L(T) \subseteq A \subseteq V(T) \cap V(H)$  are clear. We show  $V(T) \cap V(H) \subseteq L(T)$ . Assume for a contradiction that  $x \in V(T) \cap V(H)$  had degree  $d \geq 2$  in  $T$ . Let  $T_1, \dots, T_d$  be the components of  $T - x$  and for  $j \in [d]$  let  $A_j := A \cap V(T_j) \cup \{x\}$ . Since  $L(T) \subseteq A$ , every tree  $T_i$  contains some  $a \in A$  and so  $A \not\subseteq A_j$ . In particular  $|A_j| \leq k$ . Moreover  $\text{sd}_G(A_j) \leq \ell(T_j + x) < \ell(T)$ , so by our choice of  $A$  and  $T$  it follows that  $\text{sd}_G(A_j) = \text{sd}_H(A_j)$ . Therefore, for every  $j \in [d]$  there exists a connected  $S_j \subseteq H$  with  $A_j \subseteq V(S_j)$  and  $\ell(S_j) \leq \ell(T_j + x)$ . But then  $S := \bigcup_j S_j \subseteq H$  is connected, contains  $A$  and satisfies

$$\ell(S) \leq \sum_{j=1}^d \ell(S_j) \leq \sum_{j=1}^d \ell(T_j + x) = \ell(T),$$

which contradicts the fact that  $\text{sd}_H(A) > \ell(T)$  by choice of  $A$  and  $T$ .

*Claim 2:*  $E(T) \cap E(H) = \emptyset$ .

Assume for a contradiction that  $xy \in E(T) \cap E(H)$ . By Claim 1,  $x, y \in L(T)$  and so  $T$  consists only of the edge  $xy$ . But then  $T \subseteq H$  and  $\text{sd}_H(A) \leq \ell(T)$ , contrary to our choice of  $A$  and  $T$ .

*Claim 3:*  $\ell(T) < \text{sd}_H(L(T))$ .

We have  $\ell(T) = \text{sd}_G(A) < \text{sd}_H(A)$ . By Claim 1,  $A = L(T)$ .

*Claim 4:* For every  $B \subsetneq L(T)$  we have  $\text{sd}_H(B) \leq \text{sd}_T(B)$ .

Let  $B \subsetneq L(T)$  and let  $T' := T - (A \setminus B)$ . By Claim 1,  $T'$  is the tree obtained from  $T$  by chopping off all leaves not in  $B$  and so

$$\text{sd}_G(B) \leq \ell(T') < \ell(T) = \text{sd}_G(A).$$

By minimality of  $A$ , it follows that  $\text{sd}_H(B) = \text{sd}_G(B) \leq \text{sd}_T(B)$ . □

Our proofs of Theorem 1.1 and Theorem 1.2 proceed by contradiction and follow a similar outline. Let  $H \subseteq G$  be a subgraph satisfying a certain set of assumptions. The aim is to show that  $H$  is fully geodesic. Assume for a contradiction that it was not and apply Lemma 3.1 to find a shortcut tree  $T$  for  $H$ . Let  $C$  be a cycle with  $V(C) \subseteq L(T)$  and let  $W_H, W_T$  be walks traced by  $C$  in  $H$  and  $T$ , respectively. If  $|L(T)| \geq 3$ , then it follows from (2) and (SCT 4) that  $\text{len}(W_H) \leq \text{len}(W_T)$ .

Ensure that  $m_{W_T}(e) \leq 2$  for every  $e \in E(T)$  and that  $m_{W_H}(e) \geq 2$  for all  $e \in E(S)$ , where  $S \subseteq H$  is connected with  $L(T) \subseteq V(S)$ . Then

$$2 \text{sd}_H(L(T)) \leq 2 \ell(S) \leq \text{len}(W_H) \leq \text{len}(W_T) \leq 2 \ell(T),$$

which contradicts (SCT 3).

The first task is thus to determine, given a tree  $T$ , for which cycles  $C$  with  $V(C) \subseteq V(T)$  we have  $m_W(e) \leq 2$  for all  $e \in E(T)$ , where  $W$  is a walk traced by  $C$  in  $T$ . Let  $S \subseteq T$  be the Steiner tree for  $V(C)$  in  $T$ . It is clear that  $W$  does not traverse any edges  $e \in E(T) \setminus E(S)$  and  $L(S) \subseteq V(C) \subseteq V(S)$ . Hence we can always reduce to this case and may for now assume that  $S = T$  and  $L(T) \subseteq V(C)$ .

**Lemma 3.2.** *Let  $T$  be a tree,  $C$  a cycle with  $L(T) \subseteq V(C) \subseteq V(T)$  and  $W$  a walk traced by  $C$  in  $T$ . Then  $m_W(e)$  is positive and even for every  $e \in E(T)$ .*

*Proof.* Let  $e \in E(T)$  and let  $V(C) = V(C)_1 \cup V(C)_2$  be the induced bipartition. Since  $L(T) \subseteq V(C)$ , this bipartition is non-trivial. By (3),  $m_W(e)$  is the number of  $ab \in E(C)$  such that  $e \in aTb$ . By definition,  $e \in aTb$  if and only if  $a$  and  $b$  lie in different sides of the bipartition. Every cycle has a positive even number of edges across any non-trivial bipartition of its vertex-set.  $\square$

**Lemma 3.3.** *Let  $T$  be a tree,  $C$  a cycle with  $L(T) \subseteq V(C) \subseteq V(T)$ . Then*

$$2\ell(T) \leq \sum_{ab \in E(C)} d_T(a, b).$$

*Moreover, there is a cycle  $C$  with  $V(C) = L(T)$  for which equality holds.*

*Proof.* Let  $W$  be a walk traced by  $C$  in  $T$ . By Lemma 3.2, (1) and (2)

$$2\ell(T) \leq \sum_{e \in E(T)} m_W(e) \ell(e) = \text{len}(W) = \sum_{ab \in E(C)} d_T(a, b).$$

To see that equality can be attained, let  $2T$  be the multigraph obtained from  $T$  by doubling all edges. Since all degrees in  $2T$  are even, it has a Eulerian trail  $W$ , which may be considered as a walk in  $T$  with  $m_W(e) = 2$  for all  $e \in E(T)$ . This walk traverses the leaves of  $T$  in some cyclic order, which yields a cycle  $C$  with  $V(C) = L(T)$ . It is easily verified that  $W$  is traced by  $C$  in  $T$  and so

$$2\ell(T) = \sum_{e \in E(T)} m_W(e) \ell(e) = \text{len}(W) = \sum_{ab \in E(C)} d_T(a, b).$$

$\square$

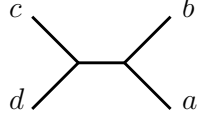


Figure 1: A tree with four leaves

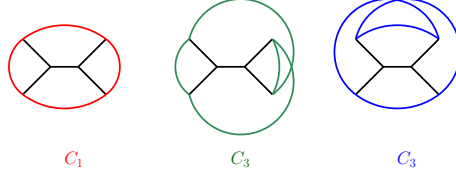


Figure 2: The three cycles on  $T$

We have now covered everything needed in the proof of Theorem 1.1, so the curious reader may skip ahead to Section 4.

In general, not every cycle  $C$  with  $V(C) = L(T)$  achieves equality in Lemma 3.3. Consider the tree  $T$  from Figure 3 and the following three cycles on  $L(T)$

$$C_1 = abcda, \quad C_2 = acdba, \quad C_3 = acbda.$$

For the first two, equality holds, but not for the third one. But how does  $C_3$  differ from the other two? It is easy to see that we can add  $C_1$  to the planar drawing of  $T$  depicted in Figure 3: There exists a planar drawing of  $T \cup C_1$  extending this particular drawing. This is not true for  $C_2$ , but it can be salvaged by exchanging the positions of  $a$  and  $b$  in Figure 3. Of course, this is merely tantamount to saying that  $T \cup C_i$  is planar for  $i \in \{1, 2\}$ .

On the other hand, it is easy to see that  $T \cup C_3$  is isomorphic to  $K_{3,3}$  and therefore non-planar.

**Lemma 3.4.** *Let  $T$  be a tree and  $C$  a cycle with  $V(C) = L(T)$ . Let  $W$  be a walk traced by  $C$  in  $T$ . The following are equivalent:*

- (a)  $T \cup C$  is planar.
- (a) For every  $e \in E(T)$ , both  $V(C)_1^e, V(C)_2^e$  are connected in  $C$ .
- (a)  $W$  traverses every edge of  $T$  precisely twice.

*Proof.* (a)  $\Rightarrow$  (b): Fix a planar drawing of  $T \cup C$ . The closed curve representing  $C$  divides the plane into two regions and the drawing of  $T$  lies in the closure of one of them. By symmetry, we may assume that it lies within the closed disk inscribed by  $C$ . Let  $A \subseteq V(C)$  be disconnected and choose  $a, b \in A$  from distinct components of  $C[A]$ .  $C$  is the disjoint union of two edge-disjoint  $a$ - $b$  paths  $S_1, S_2$  and both of them must meet  $C \setminus A$ , say  $c \in V(S_1) \setminus A$  and  $d \in V(S_2) \setminus A$ .

The curves representing  $aTb$  and  $cTd$  lie entirely within the disk and so they must cross. Since the drawing is planar,  $aTb$  and  $cTd$  have a common vertex. In particular,  $A$  cannot be the set of leaves within a component of  $T - e$  for any edge  $e \in E(T)$ .

(b)  $\Rightarrow$  (c): Let  $e \in E(T)$ . By assumption, there are precisely two edges  $f_1, f_2 \in E(C)$  between  $V(C)_1^e$  and  $V(C)_2^e$ . These edges are, by definition, the ones whose endpoints are separated in  $T$  by  $e$ . By (3),  $m_W(e) = 2$ .

(c)  $\Rightarrow$  (a): For  $ab \in E(C)$ , let  $D_{ab} := aTb + ab \subseteq T \cup C$ . The set  $\mathcal{D} := \{D_{ab} : ab \in E(C)\}$  of all these cycles is the fundamental cycle basis of  $T \cup C$  with respect to the spanning tree  $T$ . Every edge of  $C$  occurs in only one cycle of  $\mathcal{D}$ . By assumption and (3), every edge of  $T$  lies on precisely two cycles in  $\mathcal{D}$ . Covering every edge of the graph at most twice, the set  $\mathcal{D}$  is a *sparse basis* of the cycle space of  $T \cup C$ . By MacLane's Theorem,  $T \cup C$  is planar.  $\square$

## 4 Shortcut trees for trees

*Proof of Theorem 1.1.* Assume for a contradiction that  $T \subseteq G$  was not fully geodesic and let  $R \subseteq T$  be a shortcut tree for  $T$ . Let  $T' \subseteq T$  be the Steiner tree for  $L(R)$  in  $T$ . By Lemma 3.3, there is a cycle  $C$  with  $V(C) = L(R)$  such that

$$2\ell(R) = \sum_{ab \in E(C)} d_R(a, b).$$

Note that  $T'$  is 2-geodesic in  $T$  and therefore in  $G$ , so that  $d_{T'}(a, b) \leq d_R(a, b)$  for all  $ab \in E(C)$ . Since every leaf of  $T'$  lies in  $L(R) = V(C)$ , we can apply Lemma 3.3 to  $T'$  and  $C$  and conclude

$$2\ell(T') \leq \sum_{ab \in E(C)} d_{T'}(a, b) \leq \sum_{ab \in E(C)} d_R(a, b) = 2\ell(R),$$

which contradicts (SCT 3).  $\square$

## 5 Shortcut trees for cycles

By Lemma 3.1, it suffices to prove the following.

**Theorem 5.1.** *Let  $T$  be a shortcut tree for a cycle  $C$ . Then  $T \cup C$  is a subdivision of one of the five (multi-)graphs in Figure 3. In particular,  $C$  is not 6-geodesic in  $T \cup C$ .*

Theorem 5.1 is best possible in the sense that for each of the graphs in Figure 3 there exists a length-function which makes the tree inside a shortcut tree for the outer cycle, see Figure 5. These length-functions were constructed in a joint effort with Pascal Gollin and Karl Heuer in an ill-fated attempt to prove that a statement like Theorem 1.2 could not possibly be true.

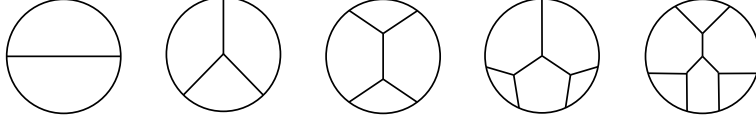


Figure 3: The five possible shortcut trees for a cycle

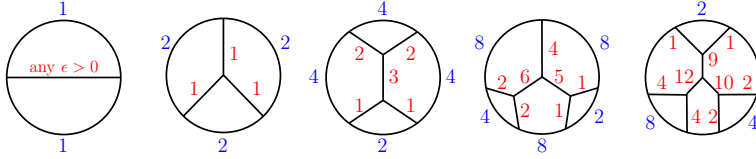


Figure 4: Shortcut trees for cycles

This section is devoted entirely to the proof of Theorem 5.1. Let  $T$  be a shortcut tree for a cycle  $C$  with length-function  $\ell : E(T \cup C) \rightarrow \mathbb{R}^+$  and let  $L := L(T)$ .

The case where  $|L| = 2$  is trivial, so we henceforth assume that  $|L| \geq 3$ . By suppressing any degree-2 vertices, we may assume without loss of generality that  $V(C) = L(T)$  and that  $T$  contains no vertices of degree 2.

**Lemma 5.2.** *Let  $T_1, T_2 \subseteq T$  be edge-disjoint trees. For  $i \in \{1, 2\}$ , let  $L_i := L \cap V(T_i)$ . If  $L = L_1 \cup L_2$  is a non-trivial bipartition of  $L$ , then both  $C[L_1], C[L_2]$  are connected.*

*Proof.* By (SCT 4) there are connected  $S_1, S_2 \subseteq C$  with  $\ell(S_i) \leq \text{sd}_T(L_i) \leq \ell(T_i)$  for  $i \in \{1, 2\}$ . Assume for a contradiction that  $C[L_1]$  was not connected. Then  $V(S_1) \cap L_2$  is non-empty and  $S_1 \cup S_2$  is connected, contains  $L$  and satisfies

$$\ell(S_1 \cup S_2) \leq \ell(S_1) + \ell(S_2) \leq \ell(T_1) + \ell(T_2) \leq \ell(T),$$

which contradicts (SCT 3).  $\square$

**Lemma 5.3.**  *$T \cup C$  is planar and 3-regular.*

*Proof.* Let  $e \in E(T)$ , let  $T_1, T_2$  be the two components of  $T - e$  and let  $L = L_1 \cup L_2$  be the induced (non-trivial) bipartition of  $L$ . By Lemma 5.2, both  $C[L_1], C[L_2]$  are connected. Therefore  $T \cup C$  is planar by Lemma 3.4.

To see that  $T \cup C$  is 3-regular, it suffices to show that no  $t \in T$  has degree greater than 3 in  $T$ . We just showed that  $T \cup C$  is planar, so fix some planar drawing of it. Suppose for a contradiction that  $t \in T$  had  $d \geq 4$  neighbors in  $T$ . In the drawing, these are arranged in some cyclic order as  $t_1, t_2, \dots, t_d$ . For  $j \in [d]$ , let  $R_j := T_j + t$ , where  $T_j$  is the component of  $T - t$  containing  $t_j$ . Let  $T_{\text{odd}}$  be the union of all  $R_j$  for odd  $j \in [d]$  and  $T_{\text{even}}$  the union of all  $R_j$  for even  $j \in [d]$ . Then  $T_{\text{odd}}, T_{\text{even}} \subseteq T$  are edge-disjoint and yield a nontrivial

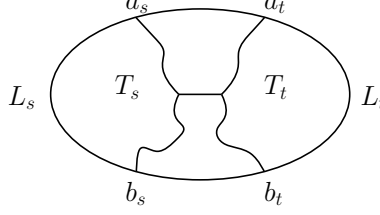


Figure 5: The setup in the proof of Lemma 5.5

bipartition  $L = L_{\text{odd}} \cup L_{\text{even}}$  of the leaves. But neither of  $C[L_{\text{odd}}], C[L_{\text{even}}]$  is connected, contrary to Lemma 5.2.  $\square$

**Lemma 5.4.** *Let  $e_0 \in E(C)$  arbitrary. Then for any two consecutive edges  $e_1, e_2$  of  $C$  we have  $\ell(e_1) + \ell(e_2) > \ell(e_0)$ . In particular  $\ell(e_0) < \ell(C)/2$ .*

*Proof.* Suppose that  $e_1, e_2 \in E(C)$  are both incident with  $x \in L$ . Let  $S \subseteq C$  be a Steiner tree for  $B := L \setminus \{x\}$  in  $C$ . By (SCT 4) and (SCT 3) we have

$$\ell(S) \leq \text{sd}_T(B) \leq \ell(T) < \text{sd}_C(L).$$

Thus  $x \notin S$  and  $E(S) = E(C) \setminus \{e_1, e_2\}$ . Thus  $P := C - e_0$  is not a Steiner tree for  $B$  and we must have  $\ell(P) > \ell(S)$ .  $\square$

Let  $t \in T$  and  $N$  its set of neighbors in  $T$ . For every  $s \in N$  the set  $L_s$  of leaves  $x$  with  $s \in tTx$  is connected in  $C$ . Each  $C[L_s]$  has two edges  $f_s^1, f_s^2 \in E(C)$  incident to it.

**Lemma 5.5.** *There is a  $t \in T$  such that for every  $s \in N$  and any  $f \in \{f_s^1, f_s^2\}$  we have  $\ell(C[L_s] + f) < \ell(C)/2$ .*

*Proof.* We construct a directed graph  $D$  with  $V(D) = V(T)$  as follows. For every  $t \in T$ , draw an arc to any  $s \in N$  for which  $\ell(C[L_s] + f_s^i) \geq \ell(C)/2$  for some  $i \in \{1, 2\}$ .

*Claim:* If  $\vec{ts} \in E(D)$ , then  $\vec{st} \notin E(D)$ .

Assume that there was an edge  $st \in E(T)$  for which both  $\vec{st}, \vec{ts} \in E(D)$ . Let  $T_s, T_t$  be the two components of  $T - st$ , where  $s \in T_s$ , and let  $L = L_s \cup L_t$  be the induced bipartition of  $L$ . By Lemma 5.2, both  $C[L_s]$  and  $C[L_t]$  are connected paths, say with endpoints  $a_s, b_s$  and  $a_t, b_t$  (possibly  $a_s = b_s$  or  $a_t = b_t$ ) so that  $a_s a_t \in E(C)$  and  $b_s b_t \in E(C)$  (see Figure 5). Without loss of generality  $\ell(a_s a_t) \leq \ell(b_s b_t)$ . Since  $\vec{ts} \in E(D)$  we have  $\ell(C[L_t] + b_s b_t) \geq \ell(C)/2$  and therefore  $C[L_s] + a_s a_t$  is a shortest  $a_t$ - $b_s$ -path in  $C$ . Similarly, it follows from  $\vec{st} \in E(D)$  that  $d_C(a_s, b_t) = \ell(C[L_t] + a_s a_t)$ .

Consider the cycle  $Q := a_t b_s a_s b_t a_t$  and let  $W_T, W_C$  be walks traced by  $Q$  in  $T$  and in  $C$ , respectively. Then  $\text{len}(W_T) \leq 2\ell(T)$ , whereas

$$\text{len}(W_C) = 2\ell(C - b_s b_t) \geq 2\text{sd}_C(L).$$

By (SCT 4) we have  $d_C(x, y) \leq d_T(x, y)$  for all  $x, y \in L$  and so  $\text{len}(W_C) \leq \text{len}(W_T)$ . But then  $\text{sd}_C(L) \leq \ell(T)$ , contrary to (SCT 3). This finishes the proof of the claim.

Since every edge of  $D$  is an orientation of an edge of  $T$  and no edge of  $T$  is oriented both ways, it follows that  $D$  has at most  $|V(T)| - 1$  edges. Since  $D$  has  $|V(T)|$  vertices, there is a  $t \in V(T)$  with no outgoing edges.  $\square$

Fix a node  $t \in T$  as guaranteed by the previous lemma. If  $t$  was a leaf with neighbor  $s$ , say, then  $\ell(f_s^1) = \ell(C) - \ell(C[L_s] + f_s^2) > \ell(C)/2$  and, symmetrically,  $\ell(f_s^2) > \ell(C)/2$ , which is impossible. Hence by Lemma 5.3,  $t$  has three neighbors  $s_1, s_2, s_3 \in T$  and we let  $L_i := C[L_{s_i}]$  and  $\ell_i := \ell(L_i)$ . There are three edges  $f_1, f_2, f_3 \in E(C) \setminus \bigcup E(L_i)$ , where  $f_1$  joins  $L_1$  and  $L_2$ ,  $f_2$  joins  $L_2$  and  $L_3$  and  $f_3$  joins  $L_3$  and  $L_1$ . Each  $L_i$  is a (possibly trivial) path whose endpoints we label  $a_i, b_i$  so that, in some orientation, the cycle is given by

$$C = a_1 L_1 b_1 + f_1 + a_2 L_2 b_2 + f_2 + a_3 L_3 b_3 + f_3.$$

Hence  $f_1 = b_1 a_2$ ,  $f_2 = b_2 a_3$  and  $f_3 = b_3 a_1$  (see Figure 5).

The fact that  $\ell_1 + \ell(f_1) \leq \ell(C)/2$  means that  $L_1 + f_1$  is a shortest  $a_1$ - $a_2$ -path in  $C$  and so  $d_C(a_1, a_2) = \ell_1 + \ell(f_1)$ . Similarly, we thus know the distance between all other pairs of vertices with just one segment  $L_i$  and one edge  $f_j$  between them.

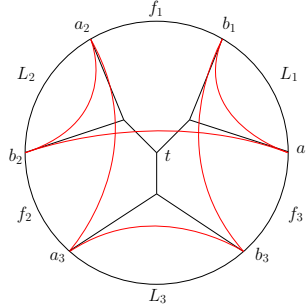


Figure 6: The cycle  $Q$

If  $|L_i| \leq 2$  for every  $i \in [3]$ , then  $T \cup C$  is a subdivision of one the graphs depicted in Figure 3 and we are done. Hence from now on we assume that at least one  $L_i$  contains at least 3 vertices.

**Lemma 5.6.** *Suppose that  $\max\{|L_s| : s \in N\} \geq 3$ . Then there is an  $s \in N$  with  $\ell(f_s^1 + C[L_s] + f_s^2) \leq \ell(C)/2$ .*

*Proof.* For  $j \in [3]$ , let  $r_j := \ell(f_{s_j}^1 + L_j + f_{s_j}^2)$ . Assume wlog that  $|L_1| \geq 3$ . Then  $L_1$  contains at least two consecutive edges, so by Lemma 5.4 we must have  $\ell_1 > \ell(f_2)$ . Therefore

$$r_2 + r_3 = \ell(C) + \ell(f_2) - \ell_1 < \ell(C),$$

so the minimum of  $r_2, r_3$  is less than  $\ell(C)/2$ .  $\square$

By the previous lemma, we may wlog assume that

$$\ell(f_2) + \ell_3 + \ell(f_3) \leq \ell(C)/2, \quad (4)$$

so that  $f_2 + L_3 + f_3$  is a shortest  $a_1$ - $b_2$ -path in  $C$ . Together with the inequalities from Lemma 5.5, this will lead to the final contradiction.

Consider the cycle  $Q = a_1 b_2 a_2 a_3 b_3 b_1 a_1$  (see Figure 5). Let  $W_T$  be a walk traced by  $Q$  in  $T$ . Every edge of  $T$  is traversed at most twice, hence

$$\sum_{ab \in E(Q)} d_T(a, b) = \ell(W_T) \leq 2\ell(T). \quad (5)$$

Let  $W_C$  be a walk traced by  $Q$  in  $C$ . Using (4) and the inequalities from Lemma 5.5, we see that

$$\begin{aligned} \ell(W_C) &= \sum_{ab \in E(Q)} d_C(a, b) = 2\ell_1 + 2\ell_2 + 2\ell_3 + 2\ell(f_2) + 2\ell(f_3) \\ &= 2\ell(C) - 2\ell(f_1). \end{aligned}$$

But by (SCT 4) we have  $d_C(a, b) \leq d_T(a, b)$  for all  $a, b \in L(T)$  and therefore  $\ell(W_C) \leq \ell(W_T)$ . Then by (5)

$$2\ell(C) - 2\ell(f_1) = \ell(W_C) \leq \ell(W_T) \leq 2\ell(T).$$

But then  $S := C - f_1$  is a connected subgraph of  $C$  with  $L(T) \subseteq V(S)$  satisfying  $\ell(S) \leq \ell(T)$ . This contradicts (SCT 3) and finishes the proof of Theorem 5.1.  $\square$

## 6 Towards a general theory

We have introduced a notion of higher geodecity based on the concept of the Steiner distance of a set of vertices. This introduces a hierarchy of properties: Every  $k$ -geodesic subgraph is, by definition, also  $m$ -geodesic for any  $m < k$ . This hierarchy is strict in the sense that for every  $k$  there are graphs  $G$  and  $H \subseteq G$  and a length-function  $\ell$  on  $G$  such that  $H$  is  $k$ -geodesic in  $G$ , but not  $(k+1)$ -geodesic. To see this, let  $G$  be a complete graph with  $V(G) = [k+1] \cup \{0\}$  and let  $H$  be the subgraph induced by  $[k+1]$ . Define  $\ell(0j) := k-1$  and  $\ell(ij) := k$  for all  $i, j \in [k+1]$ . If  $H$  was not  $k$ -geodesic, then  $G$  would contain a shortcut tree  $T$  for  $H$  with  $|L(T)| \leq k$ . Then  $T$  must be a star with center 0 and

$$\ell(T) = (k-1)|L(T)| \geq k(|L(T)| - 1).$$

But any spanning tree of  $H[L(T)]$  has length  $k(|L(T)| - 1)$  and so  $\text{sd}_H(L(T)) \leq \ell(T)$ , contrary to (SCT 3). Hence  $H$  is a  $k$ -geodesic subgraph of  $G$ . However, the star  $S$  with center 0 and  $L(S) = [k+1]$  shows that

$$\text{sd}_G(V(H)) \leq (k+1)(k-1) < k^2 = \text{sd}_H(V(H)) = k^2.$$

Theorem 1.1 and Theorem 1.2 demonstrate a rather strange phenomenon by providing situations in which this hierarchy collapses.

For a given natural number  $k \geq 2$ , let us denote by  $\mathcal{H}_k$  the class of all graphs  $H$  with the property that whenever  $G$  is a graph with  $H \subseteq G$  and  $\ell$  is a length-function on  $G$  such that  $H$  is  $k$ -geodesic, then  $H$  is also fully geodesic.

By definition, this yields an ascending sequence  $\mathcal{H}_2 \subseteq \mathcal{H}_3 \subseteq \dots$  of classes of graphs. By Theorem 1.1 all trees lie in  $\mathcal{H}_2$ . By Theorem 1.2 all cycles are contained in  $\mathcal{H}_6$ . The example above shows that  $K_{k+1} \notin \mathcal{H}_k$ .

We now describe some general properties of the class  $\mathcal{H}_k$ .

**Theorem 6.1.** *For every natural number  $k \geq 2$ , the class  $\mathcal{H}_k$  is closed under taking minors.*

To prove this, we first provide an easier characterization of the class  $\mathcal{H}_k$ .

**Proposition 6.2.** *Let  $k \geq 2$  be a natural number and  $H$  a graph. Then  $H \in \mathcal{H}_k$  if and only if every shortcut tree for  $H$  has at most  $k$  leaves.*

*Proof.* Suppose first that  $H \in \mathcal{H}_k$  and let  $T$  be a shortcut tree for  $H$ . By (SCT 3),  $H$  is not  $|L(T)|$ -geodesic in  $T \cup H$ . Let  $m$  be the minimum integer such that  $H$  is not  $m$ -geodesic in  $T \cup H$ . By Lemma 3.1,  $T \cup H$  contains a shortcut tree  $S$  with at most  $m$  leaves for  $H$ . But then by (SCT 1) and (SCT 2),  $S$  is the Steiner tree in  $T$  of  $B := L(S) \subseteq L(T)$ . If  $B \subsetneq L(T)$ , then  $\ell(S) = \text{sd}_T(B) \geq \text{sd}_H(B)$  by (SCT 4), so we must have  $B = L(T)$  and  $m \geq |L(T)|$ . Thus  $H$  is  $(|L(T)| - 1)$ -geodesic in  $T \cup H$ , but not  $|L(T)|$ -geodesic. As  $H \in \mathcal{H}_k$ , it must be that  $|L(T)| - 1 < k$ .

Suppose now that every shortcut tree for  $H$  has at most  $k$  leaves and let  $H \subseteq G$   $k$ -geodesic with respect to some length-function  $\ell : E(G) \rightarrow \mathbb{R}^+$ . If  $H$  was not fully geodesic, then  $G$  contained a shortcut tree  $T$  for  $H$ . By assumption,  $T$  has at most  $k$  leaves. But then  $\text{sd}_G(L(T)) \leq \ell(T) < \text{sd}_H(L(T))$ , so  $H$  is not  $k$ -geodesic in  $G$ .  $\square$

**Lemma 6.3.** *Let  $k \geq 2$  be a natural number and  $G$  a graph. Then  $G \in \mathcal{H}_k$  if and only if every component of  $G$  is in  $\mathcal{H}_k$ .*

*Proof.* Every shortcut tree for a component  $K$  of  $G$  becomes a shortcut tree for  $G$  by taking  $\ell(e) := 1$  for all  $e \in E(G) \setminus E(K)$ . Hence if  $G \in \mathcal{H}_k$ , then every component of  $G$  is in  $\mathcal{H}_k$  as well.

Suppose now that every component of  $G$  is in  $\mathcal{H}_k$  and that  $T$  is a shortcut tree for  $G$ . If there is a component  $K$  of  $G$  with  $L(T) \subseteq V(K)$ , then  $T$  is a shortcut tree for  $K$  and so  $|L(T)| \leq k$  by assumption. Otherwise, let  $t_1 \in L(T) \cap V(K_1)$  and  $t_2 \in L(T) \cap V(K_2)$  for distinct components  $K_1, K_2$  of  $G$ . By (SCT 4), it must be that  $L(T) = \{t_1, t_2\}$  and so  $|L(T)| = 2 \leq k$ .  $\square$

**Lemma 6.4.** *Let  $G, H$  be two graphs and let  $T$  be a shortcut tree for  $G$ . If  $G$  is a minor of  $H$ , then there is a shortcut tree  $T'$  for  $H$  which is isomorphic to  $T$ .*

*Proof.* Since  $G$  is a minor of  $H$ , there is a family of disjoint connected sets  $B_v \subseteq V(H)$ ,  $v \in V(G)$ , and an injective map  $\beta : E(G) \rightarrow E(H)$  such that for  $uv \in E(G)$ , the end vertices of  $\beta(uv) \in E(H)$  lie in  $B_u$  and  $B_v$ .

Let  $T$  be a shortcut tree for  $G$  with  $\ell : E(T \cup G) \rightarrow \mathbb{R}^+$ . By adding a small positive real number to every  $\ell(e)$ ,  $e \in E(T)$ , we may assume that the inequalities in (SCT 4) are strict, that is

$$\text{sd}_G(B) \leq \text{sd}_T(B) - \epsilon$$

for every  $B \subseteq L(T)$  with  $2 \leq |B| < |L(T)|$ , where  $\epsilon > 0$  is some constant.

Obtain the tree  $T'$  from  $T$  by replacing every  $t \in L(T)$  by an arbitrary  $x_t \in B_t$  and every  $t \in V(T) \setminus L(T)$  by a new vertex  $x_t$  not contained in  $V(H)$ , maintaining the adjacencies. It is clear by definition that  $V(T') \cap V(H) = L(T')$  and  $E(T') \cap E(H) = \emptyset$ . We now define a length-function  $\ell' : E(T' \cup H) \rightarrow \mathbb{R}^+$  as follows.

For every edge  $st \in E(T)$ , the corresponding edge  $x_s x_t \in E(T')$  receives the same length  $\ell'(x_s x_t) := \ell(st)$ . Every  $e \in E(H)$  that is contained in one of the branchsets  $B_v$  is assigned the length  $\ell'(e) := \delta$ , where  $\delta := \epsilon/|E(H)|$ . For every  $e \in E(G)$  we let  $\ell'(\beta(e)) := \ell(e)$ . To all other edges of  $H$  we assign the length  $\ell(T) + 1$ .

We now show that  $T'$  is a shortcut tree for  $H$  with the given length-function  $\ell'$ . Suppose that  $S' \subseteq H$  was a connected subgraph with  $L(T') \subseteq V(S')$  and  $\ell'(S') \leq \ell'(T')$ . By our choice of  $\ell'$ , every edge of  $S'$  must either lie in a branchset  $B_v$  or be the image under  $\beta$  of some edge of  $G$ , since otherwise  $\ell'(S') > \ell(T) = \ell'(T')$ . Let  $S \subseteq V(G)$  be the subgraph where  $v \in V(S)$  if and only if  $V(S') \cap B_v$  is non-empty and  $e \in E(S)$  if and only if  $\beta(e) \in E(S')$ . Since  $S'$  is connected, so is  $S$ : For any non-trivial bipartition  $V(S) = U \cup W$  the graph  $S'$  contains an edge between  $\bigcup_{u \in U} B_u$  and  $\bigcup_{w \in W} B_w$ , which in turn yields an edge of  $S$  between  $U$  and  $W$ . Moreover  $L(T) \subseteq V(S)$ , since  $V(S')$  contains  $x_t$  and thus meets  $B_t$  for every  $t \in L(T)$ . Finally,  $\ell(S) \leq \ell(S')$ , which contradicts our assumption that  $T$  is a shortcut tree for  $G$ .

For  $B' \subseteq L(T')$  with  $2 \leq |B'| < |L(T')|$ , let  $B := \{t \in T : x_t \in B'\}$ . By assumption, there is a connected  $S \subseteq G$  with  $B \subseteq V(S)$  and  $\ell(S) \leq \text{sd}_T(B) - \epsilon$ . Let

$$S' := \bigcup_{v \in V(S)} H[B_v] + \{\beta(e) : e \in E(S)\}.$$

For every  $x_t \in B'$  we have  $t \in B \subseteq V(S)$  and so  $x_t \in B_t \subseteq V(S')$ . Since  $S$  is connected and every  $H[B_v]$  is connected,  $S'$  is connected as well. Moreover

$$\ell'(S') \leq \delta|E(H)| + \ell(S) \leq \text{sd}_T(B) = \text{sd}_{T'}(B').$$

□

*Proof of Theorem 6.1.* Let  $H$  be a graph in  $\mathcal{H}_k$  and  $G$  a minor of  $H$ . Let  $T$  be a shortcut tree for  $G$ . By Lemma 6.4,  $H$  has a shortcut tree  $T'$  which is isomorphic to  $T$ . By Proposition 6.2 and assumption on  $H$ ,  $T$  has  $|L(T')| \leq k$  leaves. Since  $T$  was arbitrary, it follows from Proposition 6.2 that  $G \in \mathcal{H}_k$ . □

**Corollary 6.5.**  $\mathcal{H}_2$  is the class of forests.

*Proof.* By Theorem 1.1 and Lemma 6.3, every forest is in  $\mathcal{H}_2$ . On the other hand, if  $G$  contains a cycle, then it contains the triangle  $C_3$  as a minor. We saw in Section 5 that  $C_3$  has a shortcut tree with 3 leaves. By Lemma 6.4, so does  $G$  and hence  $G \notin \mathcal{H}_2$  by Proposition 6.2.  $\square$

**Corollary 6.6.** *For every natural number  $k \geq 2$  there exists an integer  $m = m(k)$  such that every graph that is not in  $\mathcal{H}_k$  has a shortcut tree with more than  $k$ , but not more than  $m$  leaves.*

*Proof.* Let  $k \geq 2$  be a natural number. By Theorem 6.1 and the Graph Minor Theorem of Robertson and Seymour [11] there is a finite set  $R$  of graphs such that for every graph  $H$  we have  $H \in \mathcal{H}_k$  if and only if  $H$  does not contain any graph in  $R$  as a minor. Let  $m(k) := \max_{G \in R} |G|$ .

Let  $H$  be a graph and suppose  $H \notin \mathcal{H}_k$ . Then  $H$  contains some  $G \in R$  as a minor. By Proposition 6.2, this graph  $G$  has a shortcut tree  $T$  with more than  $k$ , but certainly at most  $|G|$  leaves. By Lemma 6.4,  $H$  has a shortcut tree isomorphic to  $T$ .  $\square$

We remark that we do not need the full strength of the Graph Minor Theorem here: We will see in a moment that the tree-width of graphs in  $\mathcal{H}_k$  is bounded for every  $k \geq 2$ , so a simpler version of the Graph Minor Theorem can be applied, see [10]. Still, it seems that Corollary 6.6 ought to have a more elementary proof.

**Problem.** Give a direct proof of Corollary 6.6 that yields an explicit bound on  $m(k)$ . What is the smallest possible value for  $m(k)$ ?

In fact, we are not even aware of any example that shows one cannot simply take  $m(k) = k + 1$ .

Given that  $\mathcal{H}_2$  is the class of forests, it seems tempting to think of each class  $\mathcal{H}_k$  as a class of “tree-like” graphs. In fact, containment in  $\mathcal{H}_k$  is related to the tree-width of the graph, but the relation is only one-way.

**Proposition 6.7.** *For any integer  $k \geq 1$ , the graph  $K_{2,2k}$  is not in  $\mathcal{H}_{2k-1}$ .*

*Proof.* Let  $H$  be a complete bipartite graph  $V(H) = A \cup B \cup \{x, y\}$  with  $|A| = |B| = k$ , where  $uv \in E(H)$  if and only if  $u \in A \cup B$  and  $v \in \{x, y\}$  (or vice versa). We construct a shortcut tree for  $H \cong K_{2,2k}$  with  $2k$  leaves.

For  $x', y' \notin V(H)$ , let  $T$  be the tree with  $V(T) = A \cup B \cup \{x', y'\}$ , where  $x'$  is adjacent to every  $a \in A$ ,  $y'$  is adjacent to every  $b \in B$  and  $x'y' \in E(T)$ . It is clear that  $V(T) \cap V(H) = L(T)$  and  $T$  and  $H$  are edge-disjoint. We now define a length-function  $\ell : E(T \cup H) \rightarrow \mathbb{R}^+$  that turns  $T$  into a shortcut tree for  $H$ .

For all  $a \in A$  and all  $b \in B$ , let

$$\begin{aligned} \ell(ax) &= \ell(ax') = \ell(by) = \ell(by') = k - 1, \\ \ell(ay) &= \ell(ay') = \ell(bx) = \ell(bx') = k, \\ \ell(x'y') &= k - 1. \end{aligned}$$

Let  $A' \subseteq A, B' \subseteq B$ . We determine  $\text{sd}_H(A' \cup B')$ . By symmetry, it suffices to consider the case where  $|A'| \geq |B'|$ . We claim that

$$\text{sd}_H(A' \cup B') = (k-1)|A' \cup B'| + |B'|.$$

It is easy to see that  $\text{sd}_H(A' \cup B') \leq (k-1)|A'| + k|B'|$ , since  $S^* := H[A' \cup B' \cup \{x\}]$  is connected and achieves this length. Let now  $S \subseteq H$  be a tree with  $A' \cup B' \subseteq V(S)$ .

If every vertex in  $A' \cup B'$  is a leaf of  $S$ , then  $S$  can only contain one of  $x$  and  $y$ , since it does not contain a path from  $x$  to  $y$ . But then  $S$  must contain one of  $H[A' \cup B' \cup \{x\}]$  and  $H[A' \cup B' \cup \{y\}]$ , so  $\ell(S) \geq \ell(S^*)$ .

Suppose now that some  $x \in A' \cup B'$  is not a leaf of  $S$ . Then  $x$  has two incident edges, one of length  $k$  and one of length  $k-1$ . For  $s \in S$ , let  $r(s)$  be the sum of the lengths of all edges of  $S$  incident with  $s$ . Then  $\ell(s) \geq k-1$  for all  $s \in A' \cup B'$  and  $\ell(x) \geq 2k-1$ . Since  $A' \cup B'$  is independent in  $H$  (and thus in  $S$ ), it follows that

$$\begin{aligned} \ell(S) &\geq \sum_{s \in A' \cup B'} r(s) \geq |(A' \cup B') \setminus \{x\}|(k-1) + (2k-1) \\ &= |A' \cup B'|(k-1) + k \geq (k-1)|A' \cup B'| + |B'|. \end{aligned}$$

Thus our claim is proven. For  $A', B'$  as before, it is easy to see that

$$\text{sd}_T(A' \cup B') = \begin{cases} (k-1)|A' \cup B'|, & \text{if } B' = \emptyset \\ (k-1)|A' \cup B'| + k-1, & \text{otherwise.} \end{cases}$$

We thus have  $\text{sd}_T(A' \cup B') < \text{sd}_H(A' \cup B')$  if and only if  $|A'| = |B'| = k$ . Hence (SCT 3) and (SCT 4) are satisfied and  $T$  is a shortcut tree for  $H$  with  $2k$  leaves. □

Note that the graph  $K_{2,k}$  is planar and has tree-width 2. Hence there is no integer  $m$  such that all graphs of tree-width at most 2 are in  $\mathcal{H}_m$ . Using Theorem 6.1, we can turn Proposition 6.7 into a positive result, however.

**Corollary 6.8.** *For any  $k \geq 2$ , no  $G \in \mathcal{H}_k$  contains  $K_{2,k+2}$  as a minor.* □

In particular, it follows from the Grid-Minor Theorem [10] and planarity of  $K_{2,k}$  that the tree-width of graphs in  $\mathcal{H}_k$  is bounded. Bodlaender et al [4] gave a more precise bound for this special case, showing that graphs excluding  $K_{2,k}$  as a minor have tree-width at most  $2(k-1)$ .

It seems plausible that a qualitative converse to Corollary 6.8 might hold.

**Problem.** Is there a function  $q : \mathbb{N} \rightarrow \mathbb{N}$  such that every graph that does not contain  $K_{2,k}$  as a minor is contained in  $\mathcal{H}_{q(k)}$ ?

Since no subdivision of a graph  $G$  contains  $K_{2,|G|+e(G)+1}$  as a minor, a positive answer would prove the following.

**Conjecture.** For every graph  $G$  there exists an integer  $m$  such that every subdivision of  $G$  lies in  $\mathcal{H}_m$ .

## 7 Generating the cycle space

Let  $G$  be a graph with length-function  $\ell$ . It is a well-known fact (see e.g. [7, Chapter 1, exercise 37]) that the set of 2-geodesic cycles generates the cycle space of  $G$ . This extends as follows, showing that fully geodesic cycles abound.

**Proposition 7.1.** *Let  $G$  be a graph with length-function  $\ell$ . The set of fully geodesic cycles generates the cycle space of  $G$ .*

We remark, first of all, that the proof is elementary and does not rely on Theorem 1.2, but only requires Lemma 3.1 and Lemma 3.2.

Let  $\mathcal{D}$  be the set of all cycles of  $G$  which cannot be written as a 2-sum of cycles of smaller length. The following is well-known.

**Lemma 7.2.** *The cycle space of  $G$  is generated by  $\mathcal{D}$ .*

*Proof.* It suffices to show that every cycle is a 2-sum of cycles in  $\mathcal{D}$ . Assume this was not the case and let  $C \subseteq G$  be a cycle of minimum length that is not a 2-sum of cycles in  $\mathcal{D}$ . In particular,  $C \notin \mathcal{D}$  and so there are cycles  $C_1, \dots, C_k$  with  $C = C_1 \oplus \dots \oplus C_k$  and  $\ell(C_i) < \ell(C)$  for every  $i \in [k]$ . By our choice of  $C$ , every  $C_i$  can be written as a 2-sum of cycles in  $\mathcal{D}$ . But then the same is true for  $C$ , which is a contradiction.  $\square$

*Proof of Proposition 7.1.* We show that every  $C \in \mathcal{D}$  is fully geodesic. Indeed, let  $C \subseteq G$  be a cycle which is not fully geodesic and let  $T \subseteq G$  be a shortcut tree for  $C$ . There is a cycle  $D$  with  $V(D) = L(T)$  such that  $C$  is a union of edge-disjoint  $L(T)$ -paths  $P_{ab}$  joining  $a$  and  $b$  for  $ab \in E(D)$ .

For  $ab \in E(D)$  let  $C_{ab} := aTb + P_{ab}$ . Every edge of  $C$  lies in precisely one of these cycles. An edge  $e \in E(T)$  lies in  $C_{ab}$  if and only if  $e \in aTb$ . By Lemma 3.2 and (3), every  $e \in E(T)$  lies in an even number of cycles  $C_{ab}$ . Therefore  $C = \bigoplus_{ab \in E(D)} C_{ab}$ .

For every  $ab \in E(D)$ ,  $C$  contains a path  $S$  with  $E(S) = E(C) \setminus E(P_{ab})$  with  $L(T) \subseteq V(S)$ . Since  $T$  is a shortcut tree for  $C$ , it follows from (SCT 3) that

$$\ell(C_{ab}) \leq \ell(T) + \ell(P_{ab}) < \ell(S) + \ell(P_{ab}) = \ell(C).$$

In particular,  $C \notin \mathcal{D}$ .  $\square$

The fact that 2-geodesic cycles generate the cycle space has been extended to the topological cycle space of locally finite graphs by Georgakopoulos and Sprüssel [8]. Does Proposition 7.1 have a similar extension?

## References

- [1] Patrick Ali. The Steiner diameter of a graph with prescribed girth. *Discrete Math.*, 313(12):1322–1326, 2013.
- [2] Patrick Ali, Peter Dankelmann, and Simon Mukwembi. Upper bounds on the Steiner diameter of a graph. *Discrete Appl. Math.*, 160(12):1845–1850, 2012.
- [3] Patrick Ali, Simon Mukwembi, and Peter Dankelmann. Steiner diameter of 3, 4 and 5-connected maximal planar graphs. *Discrete Appl. Math.*, 179:222–228, 2014.
- [4] Hans L. Bodlaender, Jan van Leeuwen, Richard Tan, and Dimitrios M. Thilikos. On interval routing schemes and treewidth. *Inform. and Comput.*, 139(1):92–109, 1997.
- [5] J. Cáceres, A. Márquez, and M. L. Puertas. Steiner distance and convexity in graphs. *European J. Combin.*, 29(3):726–736, 2008.
- [6] Gary Chartrand, Ortrud R. Oellermann, Song Lin Tian, and Hung-Bin Zou. Steiner distance in graphs. *Časopis Pěst. Mat.*, 114(4):399–410, 1989.
- [7] R. Diestel. *Graph Theory*. Springer, 5th edition, 2017.
- [8] Agelos Georgakopoulos and Philipp Sprüssel. Geodetic topological cycles in locally finite graphs. *Electron. J. Combin.*, 16(1):Research Paper 144, 18, 2009.
- [9] Tanja Gologranc. Graphs with 4-Steiner convex balls. *Taiwanese J. Math.*, 19(5):1325–1340, 2015.
- [10] Neil Robertson and P. D. Seymour. Graph minors. V. Excluding a planar graph. *J. Combin. Theory Ser. B*, 41(1):92–114, 1986.
- [11] Neil Robertson and P. D. Seymour. Graph minors. XX. Wagner’s conjecture. *J. Combin. Theory Ser. B*, 92(2):325–357, 2004.