Structural Properties of Minimum Multi-source Multi-Sink Steiner Networks in the Euclidean Plane

Alexander Westcott[1] · Marcus Brazil[2] · Charl Ras[1]

Abstract

Given two finite sets A and B of points in the Euclidean plane, a minimum multi-source multi-sink Steiner network in the plane, or a minimum (A, B)-network, is a directed graph embedded in the plane with a dipath from every node in A to every node in B such that the total length of all arcs in the network is minimised. Such a network may contain Steiner points—nodes appearing in the solution that are neither in A nor B. We show that for any finite point sets A, B in the plane, there exists a minimum (A, B)-network that is constructible by straightedge and compass (this was claimed in a paper by Maxwell and Swanepoel, but their argument is incorrect). We use this property to formulate an algorithmic framework for exactly finding minimum (A, B)-networks in the Euclidean plane. We also present several new structural and geometric properties of minimum (A, B)-networks. In particular, we resolve a conjecture of Alfaro by proving that, for any terminal set A, adding an appropriate orientation to the edges of a minimum 2-edge-connected Steiner network on A yields a minimum (A, A)-network.

Keywords Directed Steiner networks · Steiner trees · Exact algorithm · Alfaro's conjecture · Geometric networks

1 Introduction

The Euclidean Steiner tree problem is an optimal interconnection problem, requiring a finite set of points in the plane known as terminals to be connected by a minimum
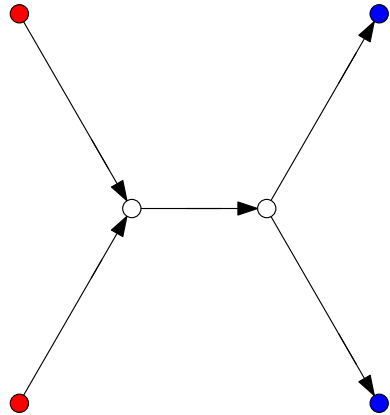
✉ Charl Ras
cjras@unimelb.edu.au

[1] School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

[2] School of Engineering, University of Melbourne, Melbourne, Australia

Fig. 1 An example of an (A, B)-network, with sources in red, and sinks in blue

length tree embedded in the plane. Solutions to this problem are known as *minimum Steiner trees*, and may include nodes that are not in the set of terminals and can lie anywhere in the plane. These additional nodes are known as *Steiner points*. Solutions to this problem are well understood, and the history of the progress in this area is detailed by Brazil et al. in [2]. Notably, in a minimum Steiner tree all Steiner points are of degree 3, with edges separated by 120°. This result and several other structural and geometric properties of optimal trees have been used to develop powerful pruning conditions, culminating in the flagship GeoSteiner exact algorithm, which is able to efficiently construct minimum Steiner trees on thousands of terminals, despite the problem being NP-hard [6].

Many types of optimal interconnection problems exist, including ones with directed connectivity requirements. One such problem is the *directed Steiner network problem* on graphs, also known as the *point to point connection problem*. Given a weighted directed graph G and a set of source sink pairs $\{(s_1, t_1), \ldots, (s_n, t_n)\}$, the problem requires the minimum weight subgraph of G containing a directed path from s_i to t_i for all i. Variations on this problem also exist where the objective is to minimise the number of nodes, or the number of edges [8]. The directed Steiner network problem has been studied primarily in terms of approximation algorithms [9], but also in the context of exact solutions when the number of source-sink pairs is fixed [8].

Not many papers have considered directed Steiner network problems in the plane. One such paper, by Maxwell and Swanepoel [13], and related papers study the following problem. Given a pair of sets A, B of points in the plane, find a shortest embedded directed network N on $A \cup B$ such there exists a directed path in N from a to b for every $a \in A$ (the set of sources) and every $b \in B$ (the set of sinks). A network satisfying this connectivity requirement is referred to as an (A, B)-network. A simple example is given in Fig. 1. For clarity, in this article we will refer to the problem of finding a minimum (A, B)-network as the *Minimum Multi-source Multi-sink Steiner Network Problem* in the plane.

Similarly to the classical Steiner tree problem, nodes of an (A, B)-network that are neither in A nor B may form part of the solution and are known as *Steiner points*.

The existence of a minimum (A, B)-network for every A, B in the plane is a nontrivial result. It is plausible that for a network with some fixed number of Steiner points, there always exists some perturbation leading to a shorter network, with the length of the networks approaching but never reaching a limit. Alternatively a similar situation could possibly arise by allowing greater numbers of Steiner points to be used: one could possibly always shorten some network by introducing more Steiner points, again approaching but never reaching a limit. However Alfaro proved that neither possibility is realised, and indeed there is some minimum (A, B)-network for every A, B in the plane [1]. In fact, similarly to how the undirected Steiner tree problem can be solved in spaces other than the Euclidean plane [11], a minimum (A, B)-network exists in any finitely compact metric space in which geodesics exist [14].

A number of structural results regarding solutions to the minimum multi-source multi-sink Steiner network problem have been established in the literature. For example, the number of Steiner points that can appear in a minimum (A, B)-network is bounded above by $13 \cdot |A \cup B|$; see [1]. In [13], Maxwell and Swanepoel provide certain necessary conditions on the local geometry and connectivity of Steiner points in minimum (A, B)-networks. In contrast to the classical Steiner tree problem, Steiner points in a minimum (A, B)-network can be of degree 3, 4, 5 or 6. It is also claimed in [13] that any minimum (A, B)-network is constructible by straightedge and compass, and consequent to this the existence of an algorithm to construct minimum (A, B)-networks is inferred. However, the proof of the claim of algebraic constructibility is flawed, as we prove in Sect. 2, and therefore the existence of an exact algorithm for constructing minimum (A, B)-networks is currently an open problem.

In this article we firstly provide a proof that given terminal sets A and B, some minimum (A, B)-network is constructible by straightedge and compass, resolving the issues presented by the incorrect claim regarding constructibility in [13]. We do this by proving that, for any A, B, there exists a minimum (A, B)-network that does not contain any cycles consisting solely of degree-3 Steiner points. Then, we use this property to present the first algorithmic framework to construct minimum (A, B)-networks. We subsequently provide several new structural properties of minimum (A, B)-networks by restricting the structure of minimum (A, B)-networks containing multiple Steiner points of degree at least 4, and improving the upper bound given in given in [1] on the number of Steiner points that can be present in a minimum (A, B)-network. Finally we disprove an open conjecture of Alfaro given in [13] by showing a minimum (A, B)-network where $A = B$ – i.e. where every terminal is both a source and a sink – can be obtained from a minimum 2-edge-connected Steiner network in the plane on terminal set A by adding appropriate orientation to edges.

The remainder of this article is structured as follows. Section 2 outlines preliminary definitions and results. Sections 3 and 4 together prove the straightedge and compass constructibility of some (A, B)-network, given A and B. An outline of an algorithm for the construction of a minimum (A, B)-network is presented in Sect. 5. Section 6 presents several structural results and Sect. 7 gives the new bound on the number of Steiner points. Finally in Sect. 8 we disprove the open conjecture given in [13].

2 Preliminaries

In this article we consider digraphs $G = (V_G, A_G)$ without self loops. We refer to elements of V_G as *nodes*, and elements of A_G *arcs*. An arc that has its head at node u is an *in-arc* of u. An arc that has its tail at u is an *out-arc* of u. We denote by $\deg^-(u)$ and $\deg^+(u)$ the number of in-arcs and out-arcs of u respectively. The *degree* of u is the total number of arcs incident to u. We define $\deg(u) := (\deg^-(u), \deg^+(u))$. We denote the arc with tail at u and head at v by (u, v).

A *geometric digraph* is an embedding of a digraph $G = (V_G, A_G)$ into a normed space. In this article, we consider the Euclidean plane. This embedding assigns each node in V_G to a point in the plane, and arcs in A_G are the straight line segments from the node at the tail to the node at the head of each arc. Throughout this article, we take $| \cdot |$ to represent length under the Euclidean norm. Given nodes (equivalently points in the plane) u and v, we define uv to be the line segment in the plane between u and v, and thus $|uv|$ is the Euclidean distance between u and v. The length of an arc $a = (u, v)$ from node u to node v is $|a| := |uv| = |vu|$.

A *dipath* between nodes u_0 and u_n in G is a finite sequence of distinct nodes $u_0, u_1 \ldots u_n$ such that $(u_i, u_{i+1}) \in A_G$ for $0 \leq i \leq n - 1$. We include a dipath from a node to itself. Given terminal sets A and B, assumed to be finite sets of distinct points in the plane (not necessarily disjoint), an (A, B)-*network* is a geometric digraph $H = (V_H, A_H)$ where $A \cup B \subseteq V_H$, and there is a dipath from every terminal in A to every terminal in B. We call this connectivity property (A, B)-*connectivity*. The nodes in A are known as *sources*, the nodes in B as *sinks*, and the nodes in $V_H \setminus A \cup B$ as *Steiner points*. The length of an (A, B)-network is defined to be $|H| := \sum_{a \in A_H} |a|$.

Definition 2.1 Given terminal sets A and B, a *minimum (A, B)-network* is an (A, B)-network H with the least value of $|H|$ amongst all (A, B)-networks.

The existence of a minimum (A, B)-network, given terminal sets A and B, was proven by Alfaro [1].

Definition 2.2 Given terminal sets A and B, the *Minimum Multi-source Multi-sink Steiner Network Problem* (MMMSNP) is to find a minimum (A, B)-network.

Throughout the remainder of this article we let A and B denote arbitrary terminal sets, and $N = (V_N, A_N)$ an arbitrary minimum (A, B)-network.

Having defined the problem, we establish some definitions that are used throughout the article. We define a *path* between nodes u_0 and u_n in N to be a finite sequence of nodes $u_0, u_1, \ldots, u_n$, such that $(u_i, u_{i+1}) \in A_N$ or $(u_{i+1}, u_i) \in A_N$, for $1 \leq i \leq n-1$. That is, a dipath in N must only use arcs in the correct orientation, and a path in N is able to use arcs regardless of direction. Note also that a dipath is a path.

We make a similar distinction between cycles depending on whether some orientation of their arcs is required or not. A *dicycle* in N is finite a sequence of distinct nodes $u_0, u_1, \ldots, u_n$, such that $(u_i, u_{i+1}) \in A_N$ for $0 \leq i \leq n - 1$, and $(u_n, u_0) \in A_N$. A *cycle* in N is finite a sequence of distinct nodes $u_0, u_1, \ldots, u_n$, such that $(u_i, u_{i+1}) \in A_N$ or $(u_{i+1}, u_i) \in A_N$, for $1 \leq i \leq n$, and $(u_n, u_0) \in A_N$ or $(u_0, u_n) \in A_N$.
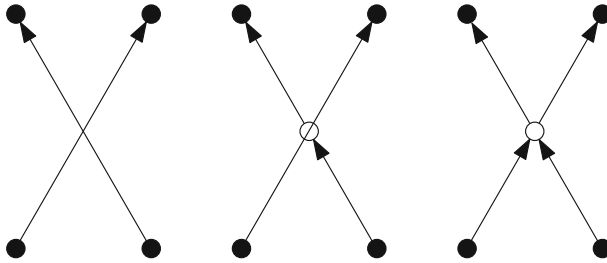
Fig. 2 Basic networks, with only the network on the right being simply embedded

A *double arc* of N is a pair of arcs (u, v) and (v, u) between two nodes u and v of N. We call the 2-cycle they form a *trivial* cycle. We note that the location of two arcs comprising a double arc coincide in the plane. The *topology* of N is the underlying digraph of N.

A minimum (A, B)-network is *simple* if for every Steiner point s in N, $\deg^-(s) \geq 1$, $\deg^+(s) \geq 1$, the degree of s is at least 3, and no degree-4 Steiner point s has two incident double arcs. Any (A, B)-network can be made simple without increasing its length so we consider only simple (A, B)-networks. This is shown for all but the last property in [13], and is true for the last property as one can replace any such s and incident arcs with a single double arc. The resultant network is no longer than N by the triangle inequality, and (A, B)-connectivity is maintained.

Definition 2.3 An (A, B)-network is *simply embedded* if every arc intersecting a node in the plane is incident to that node, and for every pair of arcs intersecting only at a single point p, there is a node at point p.

Note that this definition does not preclude arcs intersecting along a line segment as a double arc. If a minimum (A, B)-network is not simply embedded, we can make it so without increasing the length of the network, nor disrupting (A, B)-connectivity. If an arc $a = (u, v)$ intersects a node w without being incident to w, we can replace a with two arcs: (u, w) and (w, v). The sum of the lengths of these two arcs is the same as the length of a, and any dipath using a can now use the two arcs sequentially. If we have two arcs intersecting only at a single point that is not the location of a node, we place a Steiner point at that point, and apply the previous transformation. In that way, any non-simply embedded network can be transformed into a simply embedded network without increasing its length nor disrupting (A, B)-connectivity. This is illustrated in Fig. 2.

Henceforth all (A, B)-networks are assumed to be simple and simply embedded, notably including our arbitrary minimum (A, B)-network N.

We now present a definition regarding incident nodes located at the same geometrical point in the plane.

Definition 2.4 A minimum (A, B)-network is *degenerate* if it has an arc of length 0 incident to a Steiner point. We refer to such an arc as a *degenerate* arc. The *rectification* of a degenerate network is the network obtained by the following iterative process: for each Steiner point s incident to a zero length arc a of the form (u, s) or (s, u), replace
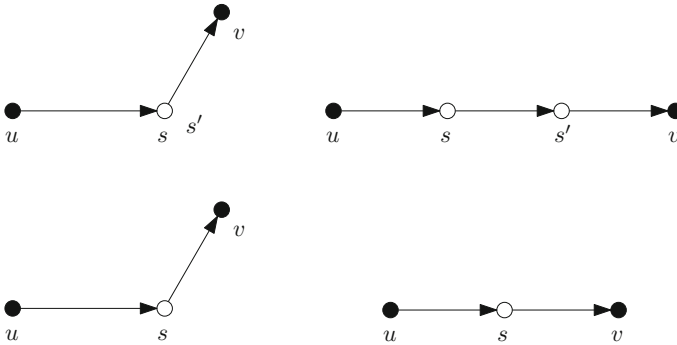
Fig. 3 Four graphs depicting a network before rectification (top two figures) and after rectification (bottom two figures). The embedding is given on the left and its topology on the right

**Fig. 4** A shortening of a network with two in-arcs meeting at less than 120° at a node
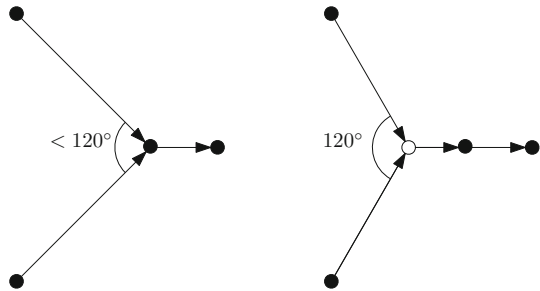
each arc of the form (v, s) with an arc (v, u), and replace each arc of the form (s, v) with an arc (u, v). Then remove s and a.

An example of rectification is given in Fig. 3. The iterative process must terminate, as (A, B)-networks are finite and an arc is removed at each iteration. A rectification of a degenerate network does not increase the length of the network as no new arcs are created, and existing arcs are either removed or remain connecting the same two points in the plane, although they may be incident to different nodes. Additionally, (A, B)-connectivity is maintained, as any dipath through a removed node can be rerouted through a node in the same location in the plane, with the equivalent arcs now incident to this node in the same orientation. We assume that all arbitrary (A, B)-networks that we reference are non-degenerate, noting that this does not exclude the explicit creation of a degenerate network.

We now outline key structural results from existing literature on the MMMSNP, and as well as relevant results from related problems. The first result is an immediate generalisation of the equivalent result for Euclidean Steiner trees [10], and is illustrated in Fig. 4.

Lemma 2.5 *In a minimum (A, B)-network, no pair of in-arcs nor pair of out-arcs meet at less than 120° at a node.*

A solution to the Euclidean Steiner tree problem is a *minimum Steiner tree*. A topology of a non-degenerate minimum Steiner tree in which every terminal has degree

(a) A Steiner point where deg(s) = (1, 2) (b) A Steiner point where deg(s) = (2, 1)
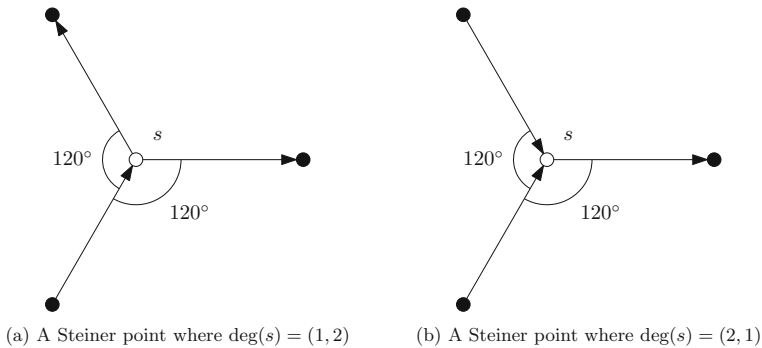
Fig. 5 Possible local geometry of a degree-3 Steiner point

1 is called a *full Steiner topology*. A non-degenerate shortest possible embedding for a full Steiner topology is called a *full Steiner tree*. Given a full Steiner topology, the corresponding full Steiner tree is unique, and is characterised by every Steiner point having degree 3, with incident edges separated at 120° (see [6]).

Lemma 2.6 *A full Steiner tree is constructible by straightedge and compass.*

This result follows immediately from the process of the Melzak-Hwang algorithm. The algorithm recursively constructs a full Steiner tree from its topology in linear time, exploiting the geometric structure of minimum Steiner trees [6].

Theorem 2.7 *(Maxwell and Swanepoel [13]) The following is a complete list of all possibilities for the local geometric structure of a Steiner point s in a minimum* (A, B)-*network in the Euclidean plane.*

1. deg(s) = (1, 2) or (2, 1). The three arcs incident to s are pairwise at 120° angles. See Fig. 5.
2. deg(s) = (2, 2). One of the following two cases:
   (a) Opposite pairs of arcs lie on two straight lines, with arcs alternating between in-arcs and out-arcs. See Fig. 6a.
   (b) Opposite pairs of arcs lie on two straight lines, arcs do not alternate between in-arcs and out-arcs, and the angles between the two in-arcs and between the two out-arcs are $\geq$ 120°. See Fig. 6b.
3. deg(s) = (2, 3), (3, 2). In the case (2, 3), the three out-arcs are pairwise at 120° degrees, and the two in-arcs lie on a straight line. The case (3, 2) is exactly opposite: the three in-arcs are pairwise at 120° degrees, and the two out-arcs lie on a straight line. See Fig. 7.
4. deg(s) = (3, 3). The in-arcs and out-arcs alternate, with consecutive arcs at 60° angles. See Fig. 8.

It has been demonstrated that cycles may be present in a minimum (A, B)-network, and that these cycles may consist solely of Steiner points, solely of terminals, or both types of nodes. This is a notable difference to the Steiner tree problem in which
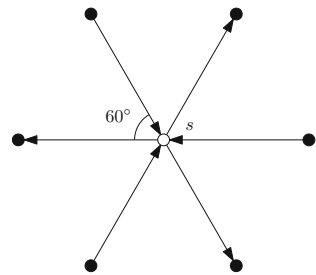
(a) $0° \leq \theta, \alpha \leq 180°$, $\alpha + \theta = 180°$　　(b) $120° \leq \theta \leq 180°$, $\alpha + \theta = 180°$
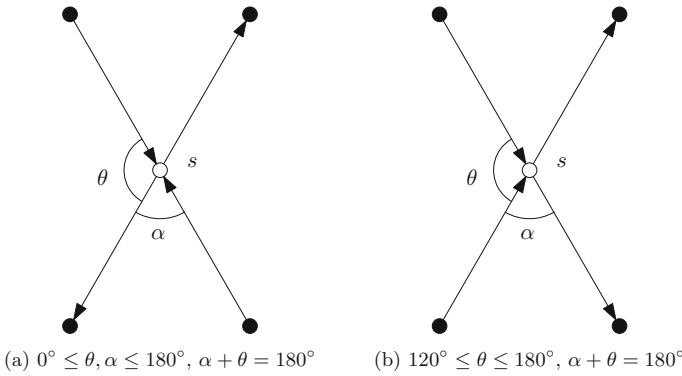
Fig. 6 Possible local geometry of a degree-4 Steiner point, where $\deg(s) = (2, 2)$

(a) A Steiner point where $\deg(s) = (2, 3)$. $0° \leq \theta \leq 120°$ (b) A Steiner point where $\deg(s) = (3, 2)$. $0° \leq \theta \leq 120°$
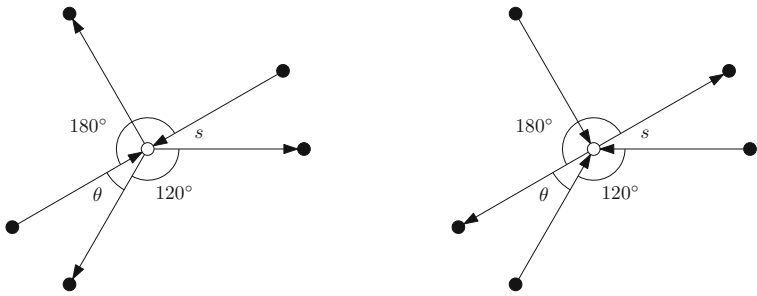
**Fig. 7** Possible local geometry of a degree-5 Steiner point

Fig. 8 Local geometry of a degree-6 Steiner point s, where $\deg(s) = (3, 3)$

solutions are always trees, and increases the number of possible optimal topologies significantly. A minimum (A, B)-network containing a cycle of Steiner points is given in Fig. 9a; a minimum (A, B)-network containing a cycle of terminals is given in Fig. 9b; and a minimum (A, B)-network containing a cycle with both a terminal and a Steiner point can be obtained by replacing a Steiner point with a terminal in Fig. 9a – minimality follows as a shorter network on that terminal set would contradict the minimality of the original, as the modified node can be restored to a Steiner point.

Maxwell and Swanepoel also claim to prove that any minimum (A, B)-network is constructible by straightedge and compass in [13], by suggesting that each such
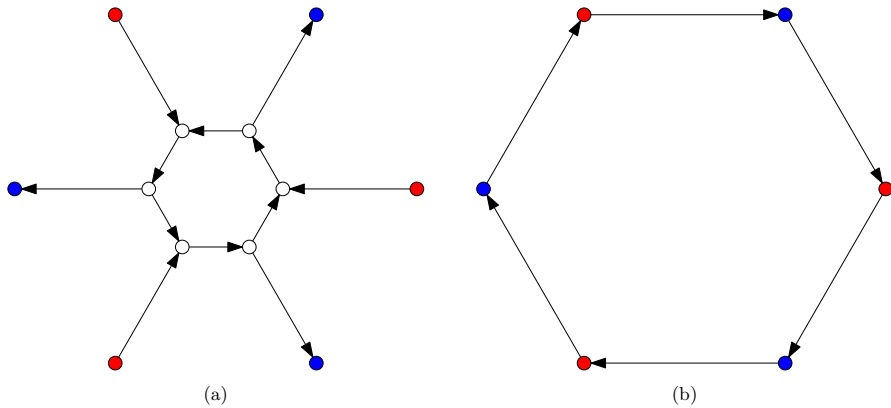
Fig. 9 Two minimum (A, B)-networks on the same set of terminals, both exhibiting cycles

network decomposes into a union of full Steiner trees, each of which is constructible by straightedge and compass by the Melzak-Hwang algorithm. However, the minimum (A, B)-network presented in Fig. 9a is a counterexample to their proof, as the cycle of degree-3 Steiner points remains a cycle after decomposing the degree-4 and degree-5 Steiner points as in their procedure. This is not a union of full Steiner trees as claimed, and exposes a flaw in the presented proof. Since a cycle of degree-3 Steiner points cannot be decomposed into a union of full Steiner trees, either some procedure to construct them from straightedge and compass must be found, or they must be excluded from necessary consideration in constructing a minimum (A, B)-network.

Given the incorrect straightedge and compass result, the subsequent claim of the existence of an algorithm for constructing minimum (A, B)-networks in [13] is invalid, and leaves the existence of such an algorithm as an open problem. In Sect. 3 we show that there exists some minimum (A, B)-network with no cycles of only degree-3 Steiner points, and we use this fact in Sect. 4 to show that some minimum (A, B)-network has a straightedge and compass construction. We then present the first exact algorithm for the construction of a minimum (A, B)-network in Sect. 5.

We now briefly discuss the complexity of the MMMSNP. If $|A| = 1$, a minimum (A, B)-network is a minimum Steiner tree [13]. Since the minimum Steiner tree problem is NP-hard [3], it follows that the minimum multi-source multi-sink Steiner network problem is NP-hard.

We also give an upper bound on the length of a minimum (A, B)-network. Given terminal sets A, B and a minimum (A, B)-network N we have that the minimum Steiner tree T on $A \cup B$, where each edge is duplicated and then directed to form a double arc, is an (A, B)-network. Therefore $2|T|$ is an upper bound for $|N|$.

In general $|T|$ is difficult to determine for large instances, so we use that $|T| \leq |\bar{T}|$, where $\bar{T}$ is a minimum spanning tree on $A \cup B$. Thus $2|\bar{T}|$ is an easily computed upper bound on the length of a minimum (A, B)-network. Both of the bounds presented are tight, shown by considering terminal sets where $A = B$ and all points of A are colinear.

This upper bound may be used in evaluating approximation algorithms for solving the problem. However in this article we primarily concern ourselves with exact solutions, and properties of minimum (A, B)-networks that may be used to create efficient exact algorithms.

3 Flexible Steiner cycles

A significant obstacle to proving the straightedge and compass constructibility of a minimum (A, B)-network is the potential for such a network to contain a cycle consisting of only degree-3 Steiner points. As discussed in Sect. 2, such cycles do not yield a decomposition into full Steiner trees allowing for the Melzak-Hwang procedure to be used. In this section we show that the Steiner points of such a cycle can always be perturbed in a manner so that length and (A, B)-connectivity are maintained until a degenerate arc is obtained. This allows us to show that for any A, B there exists a minimum (A, B)-network that does not contain such cycles. This property will be used in Sect. 4 to prove straightedge and compass constructibility of a minimum (A, B)-network.

Definition 3.1 A *degree*-3 *Steiner cycle* in N is a cycle C in N where every node on C is a degree-3 Steiner point. A node adjacent to a node on C, but not on C itself, is a *C-adjacent node*.

Definition 3.2 Let C denote a non-trivial cycle in N. The arcs composing C form a simple polygon in the plane. Let a and a' denote a pair of consecutive arcs on this polygon, both incident to a node v. The angle formed by a and a' which is interior to the polygon is called the *internal angle* of C at v. If C is a degree-3 Steiner cycle, the *cycle angle* of a C-adjacent node u is the internal angle of C at the Steiner point on C adjacent to u.

Throughout this section we let N^* denote a minimum (A, B)-network which contains at least one degree-3 Steiner cycle. We let C^* denote an arbitrary degree-3 Steiner cycle of N^*. We note that since nodes on C^* are of degree 3, there is a one to one correspondence between nodes on C^* and C^*-adjacent nodes.

An immediate consequence of Theorem 2.7 is that each internal angles of a degree-3 Steiner cycle in a minimum (A, B)-network is either 120° or 240°. Note that in N^*, a C^*-adjacent node is in the interior of the cycle if it has a cycle angle of 240°. It is in the exterior of the cycle if it has a cycle angle of 120°. We now establish a relationship between the number of each of these angles.

Lemma 3.3 *Let x and y denote the number of* 120° *and* 240° *internal angles of C^* respectively. Then $x = y + 6$.*

Proof Consider the cycle C^* as a simple polygon formed by its arcs. This polygon has $x + y$ vertices as each internal angle corresponds to a single node on C^*. The sum of internal angles of a simple polygon with k vertices is $180(k - 2)°$. Thus, we have $120x + 240y = 180(x + y - 2)$, and $x = y + 6$ follows immediately. □

We now develop the notion of *flexing C^* safely*, which represents a modification of the embedding of N^* by simultaneously translating all Steiner points on C^* (while fixing all other nodes of N^*) such that the local geometry of each Steiner point is maintained (or a degenerate arc is introduced). We show that such a flex preserves length and (A, B)-connectivity. We then use these properties to prove that there exists a minimum (A, B)-network with no degree-3 Steiner cycles.

We first define a general flex, which will not maintain the length of the network in all cases.

Definition 3.4 Let $s_1, \ldots, s_n$ denote the Steiner points on C^*. For each $i \in \{1, \ldots, n\}$, let u_i denote the C^*-adjacent node adjacent to s_i, and let ℓ_i denote the line in the plane through u_i and s_i. To *flex C^* a distance of d units*, $d \geq 0$, is to modify the embedding of N by simultaneously translating every node s_i a distance of d units along ℓ_i; for each i this translation is towards u_i if the internal angle of C^* at s_i is 120°, or away from u_i if the internal angle of C^* at s_i is 240°.

Note that the topology of N^* is maintained when flexing C^* – it is only the embedding that is modified. Crucially, however, a degenerate arc may be present after a flex. We will show that it is always possible to obtain a degenerate arc by flexing C^* appropriately, whilst maintaining length and (A, B)-connectivity.

Definition 3.5 The value δ_{C^*} is defined to be the minimum value of $d \in \mathbb{R}$, where $d \geq 0$, such that flexing C^* a distance of d units results in a degenerate arc.

For the remainder of this section we let s denote a Steiner point on C^*, s' denote the neighbour of s on C^* in the anticlockwise direction, and u and u' denote the C^*-adjacent nodes adjacent to s and s' respectively.

Lemma 3.6 *The value δ_{C^*} is finite.*

Proof By Lemma 3.3, there is at least one 120° internal angle on C^*, so we may assume that the internal angle of C^* at s is 120°. Then flexing C^* a distance of $|us|$ units results in a degenerate arc between u and s, so $\delta_{C^*} \leq |us| < \infty$.   □

Note that in the proof of the previous lemma, the local geometry of nodes on C^* may not have been maintained, for example at s' if its internal angle on C^* is also 120°, and $|u's'| < |us|$. Thus flexing in general may result in a network that is no longer a minimum (A, B)-network, and so we restrict how we may flex suitably in the following definition.

Definition 3.7 To *flex C^* safely* is to flex C^* at most δ_{C^*} units.
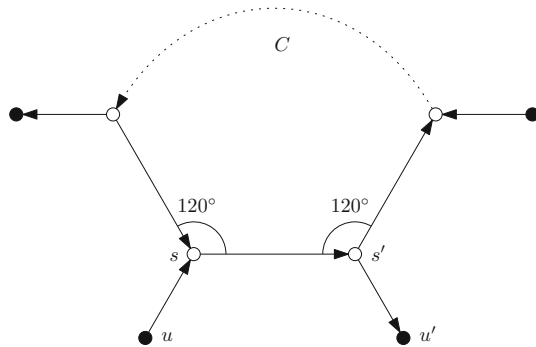
By Lemma 3.6 δ_{C^*} is finite, so it is always possible to flex C^* safely exactly δ_{C^*} units.

Definition 3.8 The *contribution* of u with respect to C^* is $|us|+|ss'|$. This is illustrated in Fig. 10.

**Fig. 10** In this network, the contribution of the C-adjacent node u is $|a_1| + |a_2|$
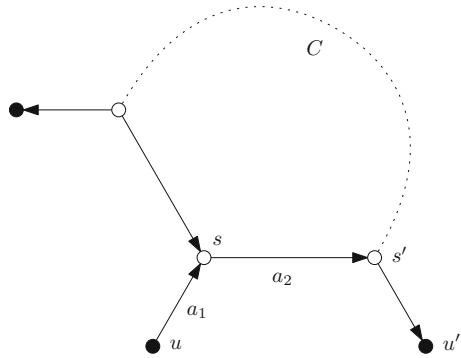
Fig. 11 The C-adjacent node u is triangular, with relevant angles displayed

We now establish a series of lemmas to show that flexing C^* safely results in a network of equal length to N^*. We first note that the length of N^* can be decomposed into three parts: the lengths of the arcs on C^*; the lengths of the arcs between C^*-adjacent nodes and C^*; and the lengths of any other arcs in the network. However, when flexing, only the arcs on C^* and the arcs between C^*-adjacent nodes and C^* are modified. Therefore, in determining the change in the length of the network after flexing, we only need to consider the sum of the lengths of these arcs in N^* and the resultant network. We note that each arc on C^* is involved in the contribution of exactly one C^*-adjacent node. Similarly, each arc between a C^*-adjacent node and C^* is involved in the contribution of exactly one C^*-adjacent node. Thus the sum of the lengths of the arcs on C^* and the arcs between C^*-adjacent nodes and C^* is equal to the sum of the contributions of all C^*-adjacent nodes. Hence, to prove that after flexing safely the length of the network is unchanged, we show that the sum of the contributions of the C^*-adjacent nodes remains the same in the resultant network. We use this to obtain our desired result: that some minimum (A, B)-network does not contain a degree-3 Steiner cycle.

Definition 3.9 If the internal angle of C^* at s is equal to the internal angle of C^* at s' then u is said to be *triangular*. This case is illustrated in Fig. 11. Otherwise, if the angles differ, u is *quadrangular*. This case is illustrated in Fig. 12.

Fig. 12 The C-adjacent node u is quadrangular, with relevant angles displayed
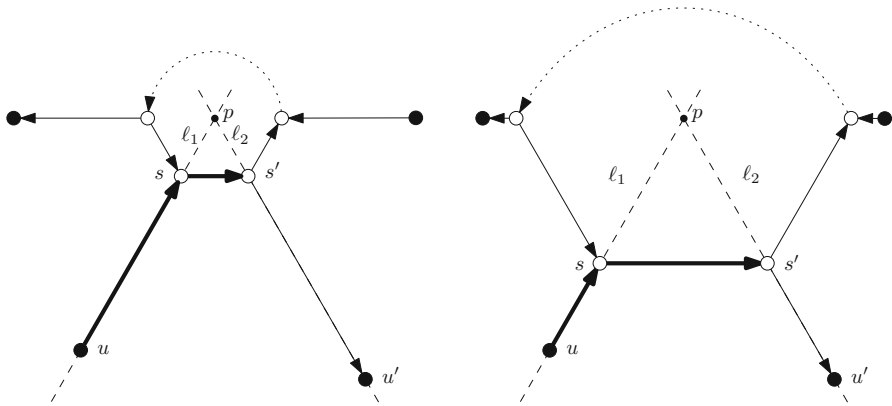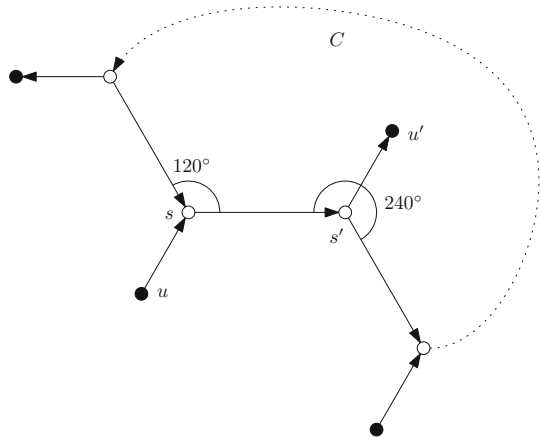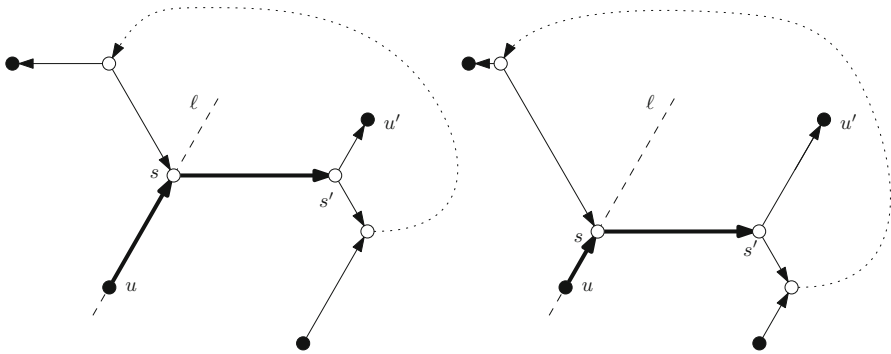

Fig. 13 A cycle before and after flexing safely, where the contribution of the triangular node u has not changed, remaining equal to the distance $|up|$. The contribution of u is the sum of the lengths of the bold arcs

Lemma 3.10 *Let u denote a triangular C^*-adjacent node. Then, flexing C^* safely does not change the contribution of u.*

Proof Let p denote the point of intersection of the line ℓ_1 passing through u and s, and the line ℓ_2 passing through u' and s'. Since u is triangular, s, s' and p are the vertices of an equilateral triangle in the plane, as in Fig. 13 (left). Hence, the contribution of u is $|up|$.

The effect of flexing C^* safely to obtain a network N' is to translate s to some point on ℓ_1 between u and s in N, and s' to some point on ℓ_2 between u' and s' in N. Since s and s' are both translated by the same distance, it follows that after translation s, s', and p still form an equilateral triangle in the embedding of N', with p at the same location in the plane. Thus flexing C^* safely does not change the contribution of u. This case is illustrated in Fig. 13.   □

Fig. 14 A cycle before and after flexing safely by d units, where the contribution of the quadrangular node u on the exterior of the cycle has decreased by d, with $|ss'|$ remaining unchanged. The contribution of u is the sum of the lengths of the bold arcs
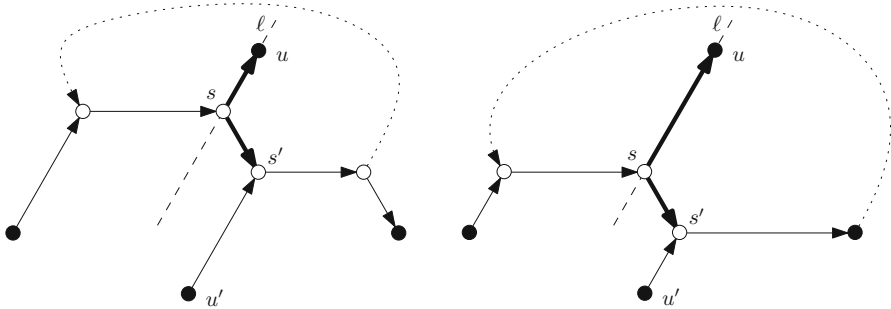
Fig. 15 A cycle before and after flexing safely by d units, where the contribution of the quadrangular node u on the interior of the cycle has increased by d, with $|ss'|$ remaining unchanged. The contribution of u is the sum of the lengths of the bold arcs. Note the right most Steiner point in the initial cycle is coincident with a node in the resultant network, with a degenerate arc between them

Lemma 3.11 *Suppose u is a quadrangular C^*-adjacent node. If u is in the exterior of C^*, then flexing C^* safely d units decreases the contribution of u by d. If u is in the interior of C^*, then flexing C^* safely d units increases the contribution of u by d.*

Proof Note that u being quadrangular implies that us and $s'u'$ are parallel. Hence, as we flex C^* safely, $|ss'|$ does not change. Therefore, the only variation to the contribution of u comes from changes in $|us|$. Let ℓ denote the line passing through u and s. Flexing C^* safely translates s either away from or towards u a distance of d units along ℓ, depending on whether u is in the interior or exterior of C^* respectively. Since $d \leq \delta_{C^*}$, s is not moved beyond u. The result follows, and is illustrated in Fig. 14 and 15. □

Lemma 3.12 *There are an equal number of quadrangular C^*-adjacent nodes in the interior of C^* and the exterior of C^*.*

Proof Suppose u is a quadrangular C^*-adjacent node. Since u is quadrangular, it is in the interior of C^* if and only if it has a cycle angle of $240°$, and u' has a cycle angle of $120°$. Similarly, u is in the exterior of C^* if and only if it has a cycle angle of $120°$, and

u' has a cycle angle of 240°. Therefore it suffices to show that in a tour of the nodes of C^*, there are an equal number of steps from a node with a corresponding interior angle of 240° to a node with a corresponding interior of 120° as there are from a node with a corresponding interior angle of 120° to a node with a corresponding interior angle of 240°.

In order to simplify our argument, we construct a new graph G and then map an anticlockwise tour of C^* to a tour in G. Let G denote the graph with exactly two vertices x, y and four arcs (x, x), (x, y), (y, x), (y, y). We think of the nodes x and y of G as corresponding to angles of 120° and 240° respectively in C^*. We start the tour of C^* at a node with a corresponding internal angle of 120°. Such a node exists as there are more 120° angles than 240° angles on C^* by Lemma 3.3. So the tour in G starts at node x.

As we tour C^* in an anticlockwise direction, if we move from a node corresponding to an internal angle of 120° to an adjacent node that also corresponds to an internal angle of 120°, then in G we move along the arc (x, x). If we move from a node corresponding to an internal angle of 120° to an adjacent node that corresponds to an internal angle of 240°, then in G we move along the arc (x, y). Analogous properties holds for the other two possibilities. Therefore, in the tour of C^*, whenever a node corresponding to an internal angle of 120° is reached then the corresponding tour in G passes through node x, and vice-versa. In G, the tour must start and end at x, as we started and finished at a 120° internal angle of C^*. Thus in the tour in G the number of times we use the arc (x, y) must be equal to the number of times we use the arc (y, x). But arcs (x, y) and (y, x) correspond exactly to quadrangular C^*-adjacent nodes in the exterior and interior of C^* respectively. So there are an equal number of quadrangular C^*-adjacent nodes in the exterior and interior of C^*. □

Lemma 3.13 *Let N' denote a network obtained from N^* by flexing C^* safely d units. Then, N' is a minimum (A, B)-network.*

Proof We have that N' is (A, B)-connected as N' has the same topology as N^*, so it remains to show $|N'| = |N^*|$. The contribution of triangular C^*-adjacent nodes has not changed from N^* to N' by Lemma 3.10. There are an equal number of quadrangular C^*-adjacent nodes in the interior and exterior of C^* by Lemma 3.12. By Lemma 3.11, the contributions in $|N'|$ of the quadrangular C^*-adjacent nodes in the interior of C^* are d units more than the contributions in N^*, and the contributions in $|N'|$ of the quadrangular C^*-adjacent nodes in the exterior of C^* are d units less than the contributions in N^*. Thus, the sum of the contributions of the C^*-adjacent nodes in N' is equal to the sum of the contributions of the C^*-adjacent nodes in N^*. Since no other part of the network changes under flexing C^*, $|N^*| = |N'|$. □

Theorem 3.14 *There exists a minimum (A, B)-network that does not contain any degree-3 Steiner cycles.*

Proof Let $N_{\min}$ denote a minimum (A, B)-network, such that it contains the minimum number of Steiner points amongst all minimum (A, B)-networks. Such an $N_{\min}$ exists as all (simple) minimum (A, B)-networks contain a finite number of Steiner points [1].

Suppose, by contradiction, that $N_{\min}$ contains a degree-3 Steiner cycle C, and let N' denote the minimum (A, B)-network obtained by flexing C safely δ_C units. By construction, N' necessarily contains a degenerate arc. Let N_0 denote the rectification of N'. This is an (A, B)-network of equal length to N' and N, as rectification preserves length and (A, B)-connectivity. However the number of Steiner points in N_0 is strictly less than the number of Steiner points in $N_{\min}$, as rectifying a degenerate network removes at least one Steiner point. This contradicts our choice of $N_{\min}$. So $N_{\min}$ cannot contain any degree-3 Steiner cycles.   □

4 Straightedge and Compass Constructibility

In this section we show that given sets of terminals A and B, there is a minimum (A, B) network which has a straightedge and compass construction. We do this by showing that minimum (A, B)-networks containing no degree-3 Steiner cycles can be decomposed into components each constructible by straightedge and compass, and that the decomposition can be reversed.

We first establish a useful proposition, allowing us to reorient the arcs in any cycle.

Definition 4.1 Let $N' = (V_{N'}, A_{N'})$ denote a (not necessarily minimum) directed network such that $A \cup B \subseteq V_{N'}$. Then we define $X(N')$ to be the set

$$X(N') = \{(u, v) \mid u \in A, v \in B, \text{ and there is a dipath from } u \text{ to } v \text{ in } N'\}$$

Recall every node is considered to have a dipath to itself, so if a node u in N' is both a source and a sink then $(u, u) \in X(N')$. We note that N' is (A, B)-connected iff $X(N') = A \times B$.

Lemma 4.2 *Let $N' = (V_{N'}, A_{N'})$ denote a (not necessarily minimum) directed network containing a cycle C' where $A \cup B \subseteq V_{N'}$. Let N'' be obtained from N' by orienting all arcs of C' to form a dicycle C''. Then $X(N') \subseteq X(N'')$.*

Proof In the case where $X(N') = \emptyset$ we are done, so we suppose that $X(N')$ is non-empty. Let $(u, v) \in X(N')$. In N'' there is a dipath between any two nodes of C'', as it is a dicycle. Now, every arc in N' is also in N'', except possibly arcs of N' which are on C'. However, if (w, x) is an arc of C' in N' but not of C'' in N'' then there is a dipath in N'' from w to x on C''. It follows that for any dipath from u to v in N' there exists a dipath from u to v in N''. So $(u, v) \in X(N'')$. Thus $X(N') \subseteq X(N'')$.   □

Proposition 4.3 *Suppose N has a cycle C. Let N' be obtained from N by orienting all arcs of C to form a dicycle. Then, N' is a minimum (A, B)-network.*

Proof Since N is an (A, B)-network, $X(N) = A \times B$. By Lemma 4.2, $A \times B = X(N) \subseteq X(N')$. So $X(N') = A \times B$. Clearly $|N'| = |N|$, hence N' is a minimum (A, B)-network.   □

Having developed this cycle orientation tool, we now proceed to show that Steiner points of degree at least four are greatly restricted in how they can appear on cycles, if they can at all.

Proposition 4.4 *Suppose N has a cycle C. Then, there is no degree-6 Steiner point on C.*

Proof Suppose, by contradiction, that C has a degree-6 Steiner point s. Then $\deg(s) = (3, 3)$, with incident in-arcs $120°$ apart, incident out-arcs $120°$ apart, and all incident arcs $60°$ apart, by Theorem 2.7. Let a_1 and a_2 denote the arcs incident to s that are on C. We have three cases: both a_1 and a_2 are in-arcs of s, both are out-arcs of s, or one is an in-arc and the other an out-arc of s.

If a_1 and a_2 are both in-arcs of s, orient C such that it is a dicycle to obtain a minimum (A, B)-network N' by Proposition 4.3. Then $\deg(s) = (2, 4)$ in N', contradicting Theorem 2.7. The case where a_1 and a_2 are both out-arcs of s is similarly contradictory.

Thus suppose without loss of generality that a_1 is an in-arc of s, and a_2 is an out-arc of s. Orienting the arcs of C to form a dicycle such that a_1 becomes an out-arc of s results in a minimum (A, B)-network N' by Proposition 4.3. But in N' the arc a_1 is $60°$ separated from another out-arc of s. This contradicts Theorem 2.7. So a degree-6 Steiner point is not on a cycle in a minimum (A, B)-network. □

**Proposition 4.5** *Suppose N has a non-trivial cycle C. Then, there is no degree-5 Steiner point on C.*

Proof Suppose, by contradiction, that C has a degree-5 Steiner point s. Then $\deg(s) = (2, 3)$ or $\deg(s) = (3, 2)$, with the two arcs of the same orientation $180°$ apart, and the three arcs of the same orientation $120°$ apart by Theorem 2.7. Call the arcs on C incident to s a_1 and a_2.

If a_1 and a_2 differ in orientation with respect to s they must not be coincident, as N is simply embedded, and C is a non-trivial cycle. Let a_3, a_4 and a_5 denote the remaining arcs incident to s. Suppose without loss of generality that a_1 and a_3 have the same orientation to s, and a_2, a_4 and a_5 the other orientation. So a_2, a_4 and a_5 are pairwise $120°$ apart, and a_1 is not $120°$ apart from both a_4 and a_5 as it is not coincident with a_2. If we orient C such that a_1 and a_2 switch orientation, we have a_1, a_4 and a_5 with the same orientation to s. But a_1 not being $120°$ apart from both a_4 and a_5 contradicts Theorem 2.7. So a_1 and a_2 cannot alternate in orientation to s.

If a_1 and a_2 have the same orientation to s, orienting C results in one of them changing direction. If they are the only two of that orientation, orienting C results in either $\deg(s) = (1, 4)$ or $\deg(s) = (4, 1)$, contradicting Theorem 2.7.

If a_1 and a_2 are two of three arcs incident to s with some orientation, we similarly consider orienting C. Call the third arc of the same orientation to s (not on C) a_3. We have that a_1, a_2 and a_3 are all $120°$ apart from one other. Orienting C results in exactly one of a_1 and a_2, say a_1, reversing orientation. Then a_2 and a_3 are the only two arcs of their orientation to s. But they are $120°$ apart, not $180°$ apart, contradicting Theorem 2.7. Therefore there can be no degree-5 Steiner point on C. □

Lemma 4.6 *Suppose s is a degree-4 Steiner point in N. Then neither a pair of in-arcs nor a pair of out-arcs incident to s are on a common cycle.*

(a) $0° < \alpha < 180°$, $\alpha + \theta = 180°$ (b) $0 < \alpha \leq 60°$, $\alpha + \theta = 180°$
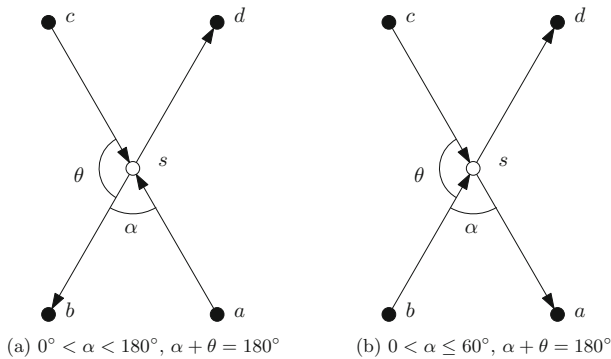
Fig. 16 Possible orientation of arcs around a degree-4 Steiner point, from Theorem 2.7. Alternating arcs are always 180° apart

Proof Suppose, by contradiction, two in-arcs (out-arcs) incident to s are on a cycle C. Then, by Proposition 4.3, orienting the arcs of C to form a dicycle gives us a Steiner point s in a minimum (A, B)-network where $\deg(s) = (1, 3)$ ($\deg(s) = (3, 1)$). This contradicts Theorem 2.7. □

Proposition 4.7 *Suppose that N has a degree-4 Steiner point s on a cycle C. Then, the internal angle of C at s is either less than or equal to* 60°*, or greater than or equal to* 300°.

Proof We suppose that the internal angle of C at s is greater than 60° and less than 300°. By Lemma 4.6, it cannot be that a pair of in-arcs nor a pair of out-arcs of s are on C.

We start with cases where in-arcs and out-arcs of s alternate around s, as in Fig. 16a. There are four relevant ways in which s can be involved in a cycle. We reference $a, b, c, d, s, \alpha, \theta$ in Fig. 16a. We define $\theta' = 360° - \theta$, and $\alpha' = 360° - \alpha$.

If the arcs (a, s) and (s, b) are on C, α or α' is the internal angle and thus $\alpha > 60°$. Then $\theta < 120°$. We orient C to form a dicycle with arcs (b, s) and (s, a), resulting in a new minimum (A, B)-network by Proposition 4.3. But (b, s) and (c, s) are both in-arcs of s, with $\angle bsc = \theta < 120°$, contradicting Theorem 2.7. Thus (a, s) and (s, b) are not on C in N.

If (c, s) and (s, b) are on C, θ or θ' is the internal angle and thus $\theta > 60°$. Then $\alpha < 120°$. We orient C to form a dicycle with arcs (b, s) and (s, c), resulting in a new minimum (A, B)-network. But now (a, s) and (b, s) are both in-arcs of s, and $\angle asb = \alpha < 120°$, contradicting Theorem 2.7. Thus (a, s) and (s, d) are not on C in N.

For the cases where (a, s) and (s, d) are on C, or where (c, s) and (s, d) are on C, equivalent arguments apply.

We now consider to the cases where in-arcs and out-arcs of s do not alternate around s, as in Fig. 16b. We have again four relevant ways in which s can be involved in a cycle, however some are contrary to our assumption that the internal angle is greater than 60° and less than 300°. We reference $a, b, c, d, s, \alpha, \theta$ in Fig. 16b, and again let $\theta' = 360° - \theta$, and $\alpha' = 360° - \alpha$.
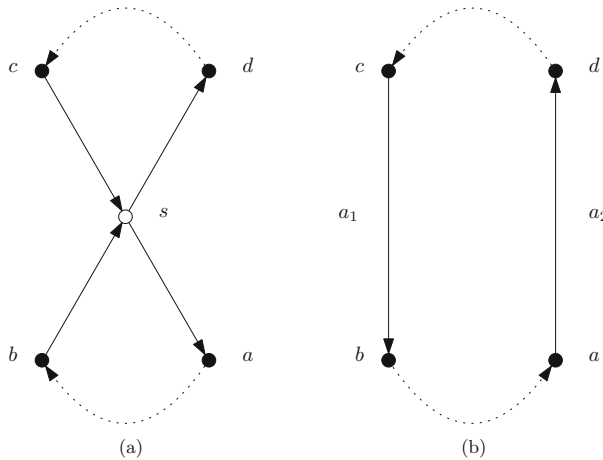
Fig. 17 An illustration of a shortening of a network containing two arc disjoint cycles sharing a degree-4 Steiner point, as used in the proof of Lemma 4.9

Since $\angle asb = \angle csd = \alpha \leq 60°$, we can immediately eliminate the cases where (s, a) and (b, s) are on C, or (c, s) and (s, d) are on C.

If (s, a) and (c, s) are on C, we orient C to form a dicycle with arcs (a, s) and (c, s). Then (a, s) and (b, s) are both in-arcs of s in the resulting minimum (A, B)-network, and $\angle asb = \alpha \leq 60° < 120°$, contradicting Theorem 2.7. A similar argument applies if (b, s) and (s, d) are on C. we orient C to form a dicycle with arcs (s, b) and (d, s) and call the resultant minimum (A, B)-network N' by Proposition 4.3. We conclude that the internal angle of C at s cannot be greater than 60° and less than 300°. □

Lemma 4.8 *Suppose that N has a degree-4 Steiner point s on a cycle C. Then, the arcs on C incident to s occur consecutively around s.*

Proof Supposing otherwise, the arcs are 180° apart by Theorem 2.7, contradicting Proposition 4.7.   □

Lemma 4.9 *No two arc disjoint cycles in N have a common degree-4 Steiner point.*

Proof Suppose that in N we have two arc disjoint cycles sharing a degree-4 Steiner point s. By Lemma 4.8 neither cycle uses non-consecutive arcs around s, so s and the cycles are as in Fig. 17a (after reorientation, if required). We refer to a, b, c, d, s as in Fig. 17, where each of a, b, c, d can be a Steiner point or a terminal. We can shorten N as in figure 17b: removing s and incident arcs, and adding arcs $a_1 = (c, b)$ and $a_2 = (a, d)$. This network is strictly shorter than N by the triangle inequality: $|cs| + |sb| > |a_1|$ and $|ds| + |sa| > |a_2|$, since $\angle cdb = \angle dsa < 180°$ by Theorem 2.7. If we then orient the new cycle in 17b, (A, B)-connectivity is maintained, as every node on the cycle has a dipath to every other node on the cycle, so dipaths in N from sources to sinks can be rerouted appropriately. Hence the minimality of N is contradicted.   □

Definition 4.10 Suppose that N contains no degree-3 Steiner cycles. The *constructible decomposition* of N is the network N_d obtained with the following adjustments, applied sequentially.

- Two arcs of all degree-5 Steiner points 180° separated are removed, and replaced by a single arc directed arbitrarily (overlapping the now degree-3 Steiner point in the plane).
- All arcs become undirected edges (with any double arc becoming a single edge).
- All degree-4 Steiner points and their incident edges are removed, with two edges added replacing the two pairs of 180° separated edges (crossing at the location of the removed Steiner point).
- All degree-6 Steiner points are duplicated, with the three former in-arcs incident only to one copy, and the three former out-arcs incident only to the other copy, so that each copy has degree 3.
- For every terminal of degree greater than one in our modified network, we make a copy of this terminal at the same location in the plane for each incident edge to the terminal, and modify the connectivity of the network so that every instance of this terminal has degree 1 (i.e. we assign to each of the instances exactly one edge incident to the original terminal).

The resulting network is the constructible decomposition of N.

We now show that the constructible decomposition of a minimum (A, B)-network is a forest, where each connected component is a full Steiner tree.

Lemma 4.11 *Suppose that N contains no degree-3 Steiner cycles. Let N_d denote the constructible decomposition of N. Then, N_d contains no cycles.*

Proof Suppose that there is a cycle C in N_d. Note that all nodes of C are degree-3 Steiner points. We have two cases: either C self-intersects in the plane, or it does not.

Suppose C self-intersects in the plane. Then, in N there is a node u of degree at least four at the point of intersection. Then u is not a terminal, as otherwise C would be disconnected in N_d. Since C is a cycle in N_d, u is on two cycles in N. These two cycles must be arc disjoint, as otherwise C would contain a repeated node, and thus would not be a cycle. By Proposition 4.4, u is not a degree-6 Steiner point, and by Proposition 4.5, u is not a degree-5 Steiner point. By Lemma 4.9, u is not a degree-4 Steiner point. Since u has at least 4 incident arcs, it is not a degree-3 Steiner point. Thus, no such u exists, and we cannot be in the first case.

Therefore, C does not self-intersect in the plane. Hence there is a cycle in N corresponding to C, possibly subdivided. Let C' denote such a cycle in N. Then C' is not a degree-3 Steiner cycle, and so contains a terminal, or a Steiner point of degree at least 4. If C' contains a terminal, it is disconnected in N_d, as the terminal is duplicated. By Proposition 4.4 a degree-6 Steiner point is not on C'. By Proposition 4.5, if a degree-5 Steiner point is on C' the cycle is trivial, and so disconnected by the decomposition of a degree-5 Steiner point. If a degree-4 Steiner point s is on C', by Proposition 4.7 the arcs of s on C' are not 180° separated, and so the cycle is disconnected in N_d. We have a contradiction, and conclude that there is no cycle in N_d.   □

The following is a corrected version of Corollary 2 in [13].

Theorem 4.12 *(Straightedge and compass) For any sets of sources and sinks A and B, there exists a minimum (A, B)-network that can be constructed from A and B by straightedge and compass.*

Proof By Theorem 3.14 there exists a minimum (A, B)-network N' containing no degree-3 Steiner cycles. We show that we can construct N' by straightedge and compass.

Let N_d denote the constructible decomposition of N'. By Lemma 4.11, N_d does not contain a cycle and hence is a forest. We note that every connected component of N_d satisfies the properties of a full Steiner tree, as every terminal is of degree 1, every Steiner point in N_d is of degree 3, and all angles around Steiner points are 120°. Some components of N_d may be self-intersecting or self-overlapping, but this does not violate the definition of a full Steiner tree.

It follows that each connected component of N_d is independently constructible by straightedge and compass, uniquely given the topology of the component, by Lemma 2.6. To obtain N' from N_d, the process of constructible decomposition is reversed: all single points of intersection between two edges in N_d are replaced by Steiner points in N'; an edge whose interior intersects with a node is replaced by two edges, increasing the degree of the node; and coincident nodes are replaced by a single terminal (if at least one of the coincident nodes is a terminal), or by a Steiner point (otherwise). Finally, edges are replaced by appropriately oriented arcs. $\square$

5 An Algorithm for Constructing Minimum (A, B)-networks

Having established a proof of the straightedge and compass constructibility of some minimum (A, B)-network, for any set of sources and sinks, we now present an exact algorithm for solving the MMMSNP. In short, the algorithm first generates all undirected full Steiner trees on two copies of the terminal set. These FSTs are overlaid upon one another, with Steiner points added at single points of intersection of edges, and with coincident nodes consolidated. The resulting undirected graph is made directed by replacing each edge with a double arc. The problem of selecting the optimal subset of arcs from this constructed directed graph is then modeled as a multi-commodity flow problem with an integer programming formulation. We now present a more detailed overview of the process.

Given sets of sources and sinks A and B, let $C = (A \cup B) \uplus (A \cup B)$, where $\uplus$ is the disjoint union, so that C is a set containing two copies of each terminal. Let S denote the set of (undirected) full Steiner trees on terminal set C. This is achieved by constructing an optimal embedding (using the Melzak procedure) of every full Steiner topology on every subset of terminals in C. Note that S may include trees self-intersecting in the plane, intersections being at a point or along a line. The disjoint union of the original terminal sets allows for two distinct terminals lying in the same location in the plane to be part of the same full Steiner tree, forming what appears to be an embedded cycle in the plane.

Let G^* denote the planar overlay of graphs in S. In other words, G^* does not have any crossing edges. Any vertex of G^* that is coincident to some terminal of a graph

in S is considered a terminal of G^*. Every other vertex of G^* is considered a Steiner point. Note that the terminal set of G^* is precisely $A \cup B$ and the set of Steiner points of G^* is a superset of the set of all Steiner points contained in a tree in S.

What follows is an integer program flow formulation for finding a minimum (A, B)-network, using G^* as input. This is a modified formulation of the multi-commodity flow formulation for Steiner trees in graphs in [6].

Consider $D = (V_D, A_D)$, the geometric digraph constructed from $G^* = (V_{G^*}, E_{G^*})$ by creating two arcs in A_D for each edge in E_{G^*} – one in each direction. Let $c(a)$ denote the length of arc $a \in A_D$. Let w denote a $|A_D|$-dimensional vector of binary variables indexed by the arcs, where $w_a = 1$ indicates that we are using arc $a \in A_D$, and $w_a = 0$ indicates we are not using arc $a \in A_D$. Let $\delta^-(v)$ denote the set of arcs with heads at node $v \in V_D$, and $\delta^+(v)$ denote the set of arcs with tails at node $v \in V_D$. For each source-sink pair (x, y) with $x \in A$, $y \in B$, $x \neq y$, we seek to send one unit of flow from x to y. We let $f^{(x,y)}$ denote the flow vector from x to y, such that $f_a^{(x,y)}$ is the flow on arc $a \in A_D$. For any $A_D' \subseteq A_D$, define $f^{(x,y)}(A_D') = \sum_{a \in A_D'} f_a^{(x,y)}$. The integer program is then:

$$\text{Minimise} \quad \sum_{a \in A_D} c(a) w_a \tag{5.1}$$

$$\text{Subject to} \quad f^{(x,y)}(\delta^+(v)) - f^{(x,y)}(\delta^-(v)) =$$
$$\begin{cases} 1, & v = x \\ -1, & v = y \\ 0, & v \in V_D \setminus \{x, y\} \end{cases} , \ v \in V_D, \ x \in A, \ y \in B, \ x \neq y \tag{5.2}$$

$$f_a^{(x,y)} \leq w_a, \ a \in A_D, \ x \in A, \ y \in B, \ x \neq y \tag{5.3}$$

$$w_a \in \{0, 1\}, \ a \in A_D \tag{5.4}$$

$$f_a^{(x,y)} \in \{0, 1\}, \ a \in A_D, \ x \in A, \ y \in B, \ x \neq y \tag{5.5}$$

A vector w minimising this integer program can be used to obtain a minimum (A, B)-network. Let $A' = \{a \in A_D \mid w_a = 1\}$. Let $V' = \{v \in V_D \mid v$ is incident to some $a \in A'\}$. Let $N' = (V', A')$. By the constraints of the integer program, N' is (A, B)-connected. By Theorem 4.12, some minimum (A, B)-network N'' decomposes into full Steiner trees constructible by straightedge and compass, and so nodes and equivalent arcs (perhaps subdivided) appearing in N'' are present in D. Thus $|N'|$ is no longer than $|N''|$. Therefore N' is a shortest (A, B)-network, albeit not necessarily simple or simply embedded.

The above formulation is not the only formulation that is possible. For example, the directed cut formulation in [6] could also be modified in a comparable manner.

We now briefly discuss the practicality of this algorithm. On a given input, it constructs every possible full Steiner tree on every subset of two copies of the terminal set. There are a superexponential number of full Steiner topologies on n terminals [6], which leads to the possibility of a superexponential number of Steiner points in G^*, and thus a superexponential number of constraints and variables in our integer

program. As such, this algorithm is currently impractical for any instance containing more than a handful of terminals.

6 Structural Properties of Minimum (A, B)-Networks

The infeasibility of considering all full Steiner topologies as in our algorithm in Sect. 5 also features in the classical Euclidean Steiner tree problem. However here is dealt with by the flagship exact algorithm (GeoSteiner [12, 15]) through the use of *pruning* conditions, which are based on strong structural and geometric constraints on optimal trees. This has resulted in an algorithm capable of solving instances of many thousands of terminals in reasonable time, notwithstanding the NP-hardness of the problem. The benefit of pruning conditions based on structural constraints has also been demonstrated for related problems such the minimum 2-edge-connected Steiner network problem in the Euclidean plane [4, 5]. This shows that research into the properties of optimal networks is crucial to efficient exact algorithm design for our problem.

In this section we therefore present a number of constraints on the topology and geometric structure of minimum (A, B)-networks. The results include that there is a minimum (A, B)-network with no degree-3 Steiner cycle, where every Steiner point is of degree at most 5 and there are at most two degree-5 Steiner points; and that no minimum (A, B)-network contains both a degree-4 Steiner point and a degree-5 Steiner point.

Lemma 6.1 *Suppose that N contains a Steiner point s of degree at least* 4. *Let a_1, a_2 denote a pair of in-arcs or a pair of out-arcs incident to s. Then a_1 and a_2 do not lie on a common cycle of N.*

Proof If s has degree 6 then a_1 and a_2 are not on a cycle by Proposition 4.4. If s has degree 5 then a_1 and a_2 cannot overlap in the plane by Theorem 2.7. Therefore, any common cycle containing a_1 and a_2 must contain more than two arcs. But no such cycle exists, by Proposition 4.5. Finally, if s has degree 4 then a_1 and a_2 are not on a common cycle by Lemma 4.6. □

Lemma 6.2 *Suppose that in N we have distinct nodes u and u' such that there are two paths from u to u', P_1 and P_2, where the first arc of each path is distinct. Then, u lies on a cycle C, where the first arc of P_1 and the first arc of P_2 are on C.*

Proof Let v denote the first node after u on P_1 that is on P_2. Such a node exists, as u' is on both paths. Since the subpaths from u to v of P_1 and P_2 respectively are disjoint except at their endpoints, it follows that there exists a cycle in N containing u, v, and the first arcs of both P_1 and P_2. □

Lemma 6.3 *Suppose that s is a Steiner point of degree at least* 4 *in N. Then each out-arc of s lies on a dipath from s to a sink, and any two dipaths from s that use distinct out-arcs of s are disjoint (except at s).*

Proof Each out-arc of s must be on some dipath to some sink, otherwise we could remove the arc and shorten the network without disconnecting any source from any

sink. Now suppose that in N there exists a node u and two dipaths from s to u, where these dipaths have distinct starting arcs which are out-arcs of s. But then, by Lemma 6.2, the two out-arcs lie on a common cycle, contradicting Lemma 6.1. The result follows. ◻

Observe that the analogous result for in-arcs of s follows from a symmetrical argument.

We define a *bridge* of a digraph D as an arc whose removal increases the number of components in the underlying graph of D.

Lemma 6.4 *Suppose that a is an arc in N that is a bridge, and K_1 and K_2 are the two disjoint connected components separated by a. Then at most one of K_1 and K_2 contains both a source and a sink. Furthermore, such a component not containing both a source and a sink is a tree.*

Proof Suppose by contradiction that both K_1 and K_2 contain both a source and a sink, and suppose without loss of generality that a is directed from K_1 to K_2 in N. Since N is (A, B)-connected, there is a dipath P from a source in K_2 to a sink in K_1, which necessarily uses a, since a is a bridge between K_1 and K_2. But a has the wrong direction for P. Thus at most one of K_1 and K_2 contains both a source and a sink. Suppose K_1 does. Let u denote the node in K_2 incident to a. If there is a cycle in K_2, we repeatedly remove arcs on cycles until we are left with a tree. We can then form an arborescence rooted at u, with directed paths from u to every terminal in K_2 (or to u from every terminal if we have sources in K_2). Applying this operation on the same arcs in N would contradict the minimality of N, and so there are no cycles in K_2. ◻

Lemma 6.5 *Suppose that N has at least two Steiner points of degree at least* 4. *Then, each such Steiner point lies on a cycle.*

Proof Let s and s' denote Steiner points of degree at least 4 in N. Suppose that s is not on a cycle. Then, every arc incident to s is a bridge. Removing all of these arcs gives at least four connected components not containing s, call four of them K_1, K_2, K_3, and K_4. Suppose without loss of generality that s' is in K_1, and that in N the arc between s and K_1 is directed from s to K_1. By Lemma 6.3, since s has at least two out-arcs, at least two of the K_i's contain sinks, so we assume without loss of generality that, besides K_1, component K_2 also contains a sink, say u. By Lemma 6.4, at most one of $K_1, \ldots, K_4$ contains both a source and a sink.

As a result of Lemma 6.3, there are two distinct sources in N with disjoint dipaths (except at s') to s'. Since these dipaths are disjoint, they do not both use the arc between s and K_1. Thus there is a source w in K_1. Since N is (A, B)-connected, there is a directed path from w in K_1 to u in K_2. But this contradicts the assumptions that that every arc incident to s is a bridge, and the arc between s and K_1 is directed from s to K_1. So s lies on a cycle. ◻

We now present a collection of structural theorems that follow from the above lemmas.

Theorem 6.6 *Suppose that there is a degree-6 Steiner point s in N. Then, every other Steiner point in N is of degree* 3.

Proof Let s' denote a distinct Steiner point in N. Suppose s' is of degree at least 4. Then, by Lemma 6.5 s lies on a cycle, contradicting Proposition 4.4. So s' is of degree 3. □

Theorem 6.7 *Suppose that N has a degree-5 Steiner point s, and another Steiner point of degree at least* 4. *Then, there is a double arc incident to s.*

Proof By Lemma 6.5, s lies on a cycle C. By Proposition 4.5, C must be a trivial cycle. So C is a double arc incident to s. □

Theorem 6.8 *Suppose that N has a degree-4 Steiner point s, and another Steiner point of degree at least* 4. *Then, the smallest angle between arcs incident to s is at most* 60°.

Proof By Lemma 6.5, s lies on a cycle C. By Proposition 4.7, the internal angle of C at s is less than or equal to 300°, or greater than or equal to 300°. Hence, the smallest angle between arcs incident to s is at most 60°. □

Theorem 6.9 *There exists a minimum (A, B)-network with no degree-3 Steiner cycles, and no degree-6 Steiner points.*

Proof By Theorem 3.14, there exists a minimum (A, B)-network N' with no degree-3 Steiner cycles. If there is no degree-6 Steiner point in N' we are done, so suppose N' has a degree-6 Steiner point s. Then by Theorem 6.6 every other Steiner point in N' is of degree 3. Let $u_0, \ldots, u_5$ denote the nodes adjacent to s in anticlockwise order, and suppose without loss of generality that the arcs between u_0, u_2, u_4 and s are in-arcs of s, and the arcs between u_1, u_3, u_5 and s are out-arcs of s, as illustrated in Fig. 18 (left). Let $d = \min\{|su_i| \mid 0 \leq i \leq 5\}$.

Denote by N'' the network obtained from N' with the following changes. Place a Steiner point s_i along the line segment su_i a distance of d away from s for $0 \leq i \leq 5$. Remove arcs $(u_0, s), (s, u_1), (u_2, s), \ldots, (s, u_5)$. Add arcs $(u_0, s_0), (s_1, u_1), (u_2, s_2) \ldots, (s_5, u_5)$, and add arcs $(s_0, s_1), \ldots, (s_4, s_5), (s_5, s_0)$. Finally remove s, and rectify the intermediate network. An example of the result of such an operation is given in Fig. 18.

For each dipath in N' from u_i to u_j via s, there is a dipath in N'' from u_i to s_i, from s_i to s_j around the dicycle, and from s_j to u_j (in the case where s_k and u_k were coincident before rectification, u_k is already on the dicycle). Hence N'' is an (A, B)-network. We also have $|N'| = |N''|$, as $|su_i| = |s_iu_i| + d$, and $d = |s_is_{i+1 \pmod 6}|$ as arcs incident to s are 60° separated by Theorem 2.7. Hence, N'' is a minimum (A, B)-network.

Finally, there are no degree-6 Steiner points in N'', as we removed the only degree-6 Steiner point in N', and at most two Steiner points of degree 3 were coincident in the plane before rectification (in the case where $\arg\min\{|su_i| \mid 0 \leq i \leq 5\}$ contains a Steiner point). In this case any coincident Steiner points are adjacent to each other, meaning that rectification could not have resulted in a degree-6 Steiner point. There

are also no degree-3 Steiner cycles in N'', as there are no degree-3 Steiner cycles in N' and rectification destroys any such cycle that we constructed, due to the necessary occurrence of coincident nodes $\arg\min\{|su_i| \mid 0 \le i \le 5\}$ and the corresponding s_i. Thus, the theorem follows.   □

We now extend some of these results by considering properties of double arcs in minimum (A, B)-networks. This further restricts the structure of minimum (A, B)-networks containing multiple Steiner points of degree at least 4.

Lemma 6.10 *Suppose that two cycles C_1 and C_2 in N share a node u, and share no arc incident to u. Then, the two cycles are trivial, separated by $180°$ at u, and u is a degree-4 terminal.*

Proof Let a_1 and a_2 denote the two arcs on C_1 incident to u, and a_3 and a_4 the two arcs on C_2 incident to u. Suppose that either at least one of C_1 and C_2 is non-trivial, or that both are trivial but separated by less than $180°$.

Then, without loss of generality a_1 and a_3 are separated by less than $180°$ at u. Let v denote the end node of a_1 that is not u, and w the end node of a_3 that is not u. Remove a_1 and a_3 from N, add an arc a_5 from w to v. The consecutive traversal of a_5, the path from v to u (on C_1 in N, not using a_1), and the path from u to w (on C_2 in N, not using a_3) gives a cycle C' containing a_5, a_4 and a_2 (either directly, or as a subpath of what is obtained if C_1 and C_2 are not disjoint). Orient C' to form a dicycle in a network N'. Since $\angle vuw < 180°$, we have $|N'| < |N|$ by the triangle inequality. We also note that N' is (A, B)-connected, as any dipath in N using a_1 or a_3 can be modified to use the dicycle formed in the construction of N'. We have thus contradicted the minimality of N, and conclude that C_1 and C_2 are both trivial, and separated by $180°$. An example of a shortening where both C_1 and C_2 are non-trivial is given in Fig. 19.

Furthermore, Lemma 2.5 implies that no other arc is incident to u, as otherwise it would be separated by less than $120°$ from an arc with the same orientation with respect to u on one of the double arcs. Hence, u has degree 4. By Theorem 2.7, and because N is simple, u must be a terminal.   □

We now present the remaining structural theorems in this section that follow from the above results.

Theorem 6.11 *Suppose that there is a double arc incident to some node in N. Then there is a sequence of distinct colinear nodes $u_0, u_1, \ldots, u_k$ such that $(u_i, u_{i+1}) \in A_N$ and $(u_{i+1}, u_i) \in A_N$ for $0 \le i \le k-1$. Furthermore, N contains no additional cycles – neither trivial nor non-trivial.*

Proof Let u denote a node with an incident double arc in N. We have two cases, either there are multiple double arcs incident to u, or just one. In the first case, by Lemma 6.10 there are exactly two double arcs incident to u $180°$ apart, and no other incident arcs. Since N is finite, there exists a maximal sequence of distinct colinear nodes $u_0, u_1, \ldots, u_k$, such that u_i and u_{i+1}, where $0 \le i \le k-1$, have a double arc between them, and $u_1, \ldots, u_{k-1}$ are degree-4 terminals. It remains to show that there are no other cycles in N – trivial or non-trivial.
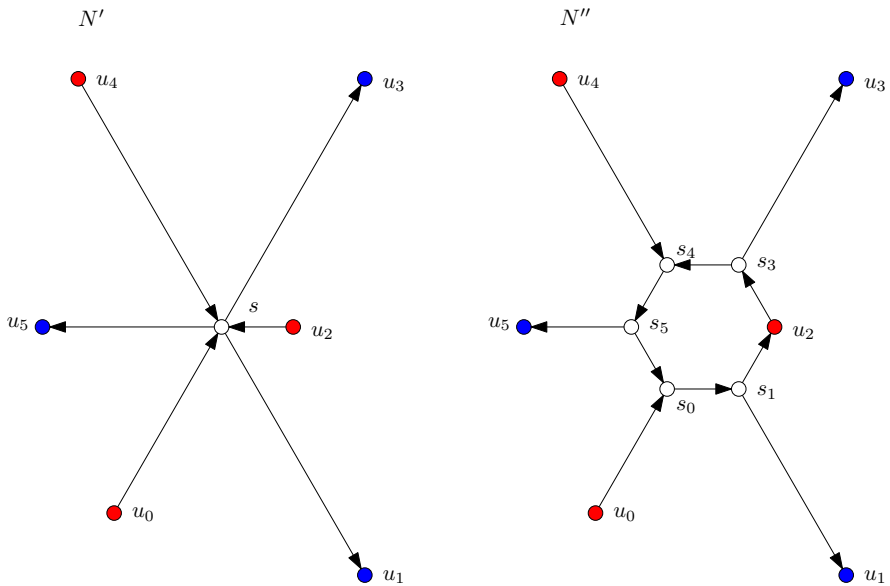
Fig. 18 An example of the operation on a network containing a degree-6 Steiner point used in the proof of Theorem 6.9
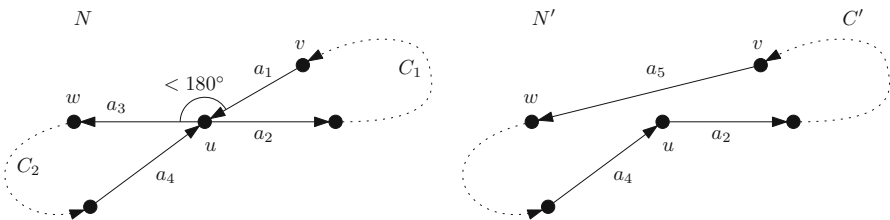
Fig. 19 A shortening of a network N containing two arc disjoint non-trivial cycles sharing a common node u

By Lemma 6.10, u_0 is on no cycle other than the trivial cycle between it and u_1. Suppose a is an arc incident to u_0 distinct from the arcs forming the double arc. Then a is a bridge in N. Let K_1 denote the connected component containing u_0 after removing a from N, and K_2 the other connected component. The arc (u_1, u_0) must be on a dipath in N from some source in K_1, and the arc (u_0, u_1) must be on a dipath in N to some sink in K_1, so K_1 contains both a source and a sink. Thus by Lemma 6.4 K_2 is a tree.

Since a was arbitrary, and by the symmetry of u_0 and u_k, we conclude that there are no cycles in N other than the trivial cycles between u_i and u_{i+1}, where $0 \leq i \leq k-1$.
□

Theorem 6.12 *There are at most two degree-5 Steiner points in N.*

Proof By Lemma 6.5, each degree-5 Steiner point in N lies on a cycle. Thus, by Proposition 4.5 each degree-5 Steiner point in N has an incident double arc. By Theorem 2.7, each degree-5 Steiner point in N has at most one incident double arc.

But by Theorem 6.11 at most two nodes in N have exactly one incident double arc. Therefore there are at most two degree-5 Steiner points in N. □

Theorem 6.13 *There is not both a degree-4 and a degree-5 Steiner point in N.*

Proof Suppose that N contains a degree-4 Steiner point s and a degree-5 Steiner point s'. By Lemma 6.5, both s and s' lie on a cycle. Thus, by Proposition 4.5 s' has an incident double arc. Then, by Theorem 6.11 there is no non-trivial cycle in N. But s is on a non-trivial cycle as it is on a cycle, and N is both simple and simply embedded. Therefore, there is not both a degree-4 and a degree-5 Steiner point in N. □

7 An Improved Bound on the Number of Steiner Points

A tightened upper bound on the number of Steiner points in a minimum (A, B)-network provides an additional tool for restricting the topologies of optimal solutions. In this section we improve the bound provided by Alfaro in [1] by reducing it from $13 \cdot |A \cup B|$ to $2 \cdot |A \cup B| - 2$, and in doing so we also provide further structural properties of minimum (A, B)-networks.

Definition 7.1 Let G denote a directed planar graph. A *region* of G is the closure of a face of G. A node in G that lies in exactly three regions of G is called *tri-regional*.

Lemma 7.2 *Suppose that u is a node in at least three regions of N. Then, u is tri-regional and has exactly three incident arcs $120°$ apart from one another, each pair of which lies on a common cycle.*

Proof We first show that every region of N has a single cycle in N composing its boundary. Suppose otherwise, and let R denote such a region of N. Note that any tree structures defining a face of N are irrelevant to the boundary of a region, due to the taking of the closure of a face. By assumption N has more than one region, so R has a boundary. So there are at least two arc disjoint non-trivial cycles composing the boundary of R. By Lemma 6.10, these two cycles do not share a node. There is at most one path (up to double arcs) between these cycles in N, otherwise R is not a region. By Lemma 6.4, this path cannot contain a bridge, and so consists of double arcs. However this contradicts Theorem 6.11. Thus, every region of N has a single cycle in N composing its boundary.

We note that if two arcs incident to u are on the boundary of the same two distinct regions, they must be the only arcs incident to u on a cycle. So if u is in at least three regions, no pair of arcs incident to u bordering regions border the same two regions. So u is on at least three cycles defining regions in N, each of which uses a pair of arcs incident to u which differ from the pair for each other region-defining cycle. Each of these cycles is non-trivial, so by Lemma 6.10, they pairwise share an arc incident to u. Hence, there are at most three such cycles. Thus, u is in exactly three regions, and so is tri-regional.

Furthermore, since each of the three region-defining cycles identified can independently be reoriented in either direction by Proposition 4.3, u must have exactly three incident arcs $120°$ separated, as otherwise a cycle could be reoriented to make two arcs, separated by less than $120°$, both in-arcs or both out-arcs of u. □

For clarity, we give the following definition of a chord path of a cycle in N.

Definition 7.3 A *chord path* of a cycle C of N is a path between two distinct nodes on C, disjoint from C except at the endpoints.

Whilst the following proposition is an intermediate result for our eventual bound, it may also have utility in increasing the efficiency of an algorithm solving MMMSNP, in the vein of the main results in Sect. 6.

**Proposition 7.4** *Suppose that P is a chord path of a cycle C of N. Then, P has at least three arcs.*

Proof Suppose first that P is a single arc (u, v) between nodes u and v on C. By Proposition 4.3, we can orient C to be a dicycle, and obtain a minimum (A, B)-network N'. Then, we have a dipath from u to v along C. So we can remove the arc (u, v) and maintain (A, B)-connectivity, contradicting the minimality of N' and N. So P is not a single arc.

Suppose now that P has two arcs. Let u and v be on both C and P, and w the other node on P. Suppose that both arcs of P are in-arcs of w or out-arcs of w. Then, after orienting C, we could remove one of these arcs without destroying (A, B)-connectivity. Thus P must be a dipath. So suppose without loss of generality that P is comprised of arcs $a_1 = (u, w)$ and $a_2 = (w, v)$. As before, we orient C to be a dicycle to obtain a minimum (A, B)-network N'. We have that $|a_1| = |a_2|$, for otherwise we could remove the longer arc and make the shorter a double arc, which shortens the network whilst maintaining (A, B)-connectivity with the dipath around C. We now remove a_1 and add an arc (v, w) to obtain the graph N'', with length and (A, B)-connectivity maintained. But now u has an incident double arc and is on a non-trivial cycle, contradicting Lemma 6.10. Hence P has at least three arcs. □

Lemma 7.5 *Suppose that P is a chord path of a cycle of N. Then there are two consecutive non-tri-regional nodes on P.*

Proof Let Q denote the subpath of P with the least number of arcs that is itself a chord path of some cycle C of N. By Proposition 7.4, Q has at least three arcs. Let u denote the first node of Q. If the two nodes after u on Q are non-tri-regional we are done, so suppose one is tri-regional, and call it v. Then, by Lemma 7.2 v is of degree 3, and each pair of arcs incident to v is on a common cycle. Let a_1 denote the arc incident to v that is not on P, and a_2 and a_3 denote the other two arcs incident to v. Since a_1 and a_2 are on a common cycle, there is either a path P_C from v to a node on C without passing through a node on Q, or there is a path P_Q from v to a node on Q without passing through a node on C. The case where v is immediately after u on Q is illustrated in Fig. 20.

In the first case, let w denote the node on C. Then, the path from u to v along Q is a chord path to the cycle that is the consecutive traversal of the path from v to the end of Q at node x, the path along C from x to w via u, and P_C from w to v, as illustrated in Fig. 20a. This chord path is shorter than Q as v is not an endpoint of Q, so the minimality of Q is contradicted.
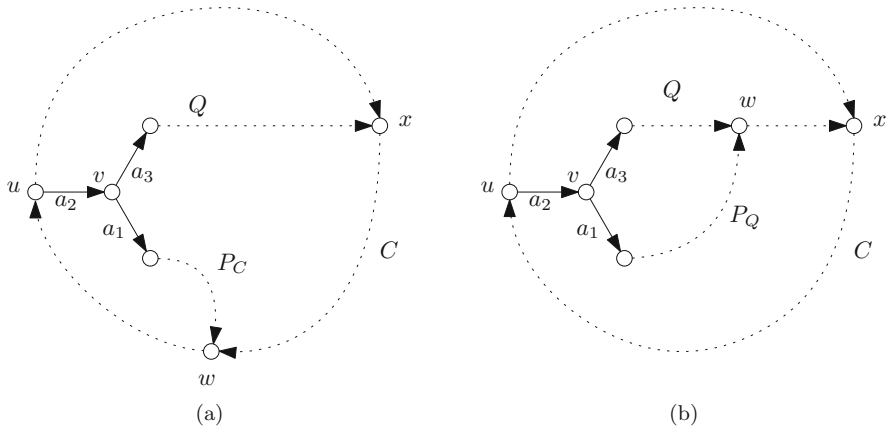
Fig. 20 Possible connectivity of a tri-regional node v on a chord path Q (running from u to x using arcs a_2 and a_3) to a cycle C, as referenced in the proof of Lemma 7.5. In the network on the left, the subpath of Q from u to v is itself a chord path to a cycle. In the network on the right, the subpath of Q from v to w is itself a chord path to a cycle

In the second case, let w denote the node on Q. Then, the path from v to w along Q is a chord path to the cycle that is the consecutive traversal of the path from w to the end of Q at node x, the path along C from x to u (anticlockwise say), the path from u to v along Q, and P_Q from v to w, as illustrated in Fig. 20b. This chord path is shorter than Q as v is not an endpoint of Q, so again the minimality of Q is contradicted. We conclude that the first two nodes of Q after u are non-tri-regional. Then, these nodes are also on P, as Q is a subpath of P. So there are two consecutive non-tri-regional nodes on P. □

Lemma 7.6 *Suppose that C is a cycle of N. Then, there are two consecutive non-tri-regional nodes on C.*

Proof If no node on C is tri-regional, we are done. Otherwise, suppose node u on C is tri-regional. Let a denote an arc on C incident to u. Then, by Lemma 7.2, u has an incident arc a' that is not on C, but is on some cycle C' such that a is also on C'. Consider the path P along C' beginning at u along arc a', until the first node that is on C. Call this node v. On C we have two paths from v to u: P' and P''. Then, P'' is a chord-path to the cycle that is the consecutive traversal of P and P'. By Lemma 7.5, there are two consecutive non-tri-regional nodes on P''. So there are two consecutive non-tri-regional nodes on C. □

Our method for bounding the number of Steiner points in N is to remove arcs from N in such a way as to leave behind a tree where every leaf is a terminal of N. We associate each removed arc with two distinct terminals of N, such that no pair of removed arcs share an associated terminal. This bounds above the number of arcs removed to form a tree. Further properties of trees allow us to bound the number of Steiner points in the tree, thus bounding the number of Steiner points in N.

Definition 7.7 Let N denote a minimum (A, B)-network, and let r denote the number of regions of N. Define $T_r(N)$ to be the set $\{N\}$, and for $1 \leq i \leq r - 1$ define $T_i(N)$ to be the set of graphs obtained from any $G \in T_{i+1}(N)$ by removing an arc a that lies on the boundary of two regions of G, where neither of the endpoints of a is tri-regional in N. Define such an arc of $G \in T_{i+1}(N)$, should one exist, to be *safely disposable*.

We note that removing an arc on the boundary of two regions of a planar graph decreases the number of regions in that graph by one, so elements of $T_k(N)$ – should they exist – have k regions.

We show that in fact we always have $T_1(N) \neq \emptyset$, and that elements of $T_1(N)$ are trees where every leaf is a terminal of N. The existence of such a tree is non-trivial, as a naive attempt to create a tree could remove two arcs incident to a tri-regional Steiner point, leaving behind a leaf that is a Steiner point.

Lemma 7.8 *Let r denote the number of regions of N. Then for $1 \leq i \leq r$, $T_i(N) \neq \emptyset$.*

Proof If $r = 1$ we are done, so suppose that there are multiple regions in N. Since $T_r(N) = \{N\}$, $T_r(N) \neq \emptyset$. Suppose $T_{i+1} \neq \emptyset$ for some $1 \leq i \leq r - 1$. We show $T_i \neq \emptyset$.

Let $G \in T_{i+1}$. Then G has more than one region, as $i + 1 > 1$. So there is a cycle C in G. Since C is also a cycle in N, by Lemma 7.6 there are two consecutive non-tri-regional nodes in N on C. So the arc a between them is safely disposable in G. Thus, by removing a from G we obtain an element of $T_i(N)$. ☐

Lemma 7.9 *If $T \in T_1(N)$ and u is a node in N then at most one arc incident to u in N is not present in T.*

Proof Suppose that $T \in T_1(N)$. Let u denote a node in N. Suppose in $G \in T_i(N)$, $1 < i \leq r$, u has an incident arc a that is safely disposable. The node u is not tri-regional in N so, by Lemma 7.2, u is in at most two regions of N. However a lies on the boundary of two regions of G, so u is in exactly two regions of N. If a were to be removed from G to form G', the node u would lie in exactly one region of G'. No arc incident to u is safely disposable in G', since if u is in exactly one region no arc incident to u can be on the boundary of two regions. This remains true in graphs obtained by removing additional arcs from G', as no arc can be removed such that a node in one region is subsequently in two regions after the removal of that arc. Hence, in T, at most one arc is incident to u in N has been removed. ☐

Lemma 7.10 *If $T \in T_1(N)$, T is a tree where every leaf of T is a terminal of N.*

Proof Suppose that $T \in T_1(N)$. Every arc of N not present T is on a cycle in the network it is removed from, so T is connected. Since T has one region, it is thus a tree. By Lemma 7.9, every node in T has at most one additional incident arc in N. As every Steiner point in N has degree at least three, every Steiner point has degree at least two in T, and so no Steiner point in N is a leaf of T. ☐

Theorem 7.11 *Let $n = |A \cup B|$. Then, the number of Steiner points in N is bounded above by $2n - 2$.*

Proof Let r denote the number of regions of N. By Lemma 7.8, $T_1(N) \neq \emptyset$, so let $T \in T_1(N)$. We seek to associate each arc of N removed in the process of forming T with two distinct terminals, such that no pair of removed arcs shares an associated terminal. We do this by associating each node incident to a removed arc with a single unique terminal.

Let a denote an arc of N not present in T, and let u denote a node incident to a in N. Then by the construction of T the node u is non-tri-regional in N, but u is on a non-trivial cycle in N as a is on the boundary of two regions of N. By Lemma 7.9, a is the only arc incident to u in N that is not in T.

If u is a terminal, we associate u to itself. If u is not a terminal, it is a Steiner point and so has degree at least three in N. It thus has an incident arc a' that is not on a non-trivial cycle in N, as u is non-tri-regional in N. Furthermore, a' is not on a trivial cycle in N by Theorem 6.11. Suppose that a' is an out-arc of u (the other case is a symmetrical argument). Then, a' is on a dipath in N from some source to some sink v. We let v denote the terminal associated to u. Since a' is not on a cycle it is a bridge. Let K and K' denote the connected components containing v and u respectively after removing the bridge a' from N. Since u is on a cycle, there must be both a source and a sink in K. Then by Lemma 6.4, every terminal in K' is of the same type, and so K' is a tree. Notably, every node on the dipath from u to v (with the exception of u) is not on a cycle in N. Hence, no other node incident to an arc of N not present in T can be associated with v by this process.

Thus, we have associated two distinct terminals with every arc of N removed to form T, each of which is not associated in the same manner with any other such arc. So at most $\frac{n}{2}$ arcs of N are removed to form T.

By Lemma 7.10, we have that T is a tree with at most n leaves, as every leaf of T is a terminal. Hence, there are at most $n - 2$ Steiner points of degree at least 3 in T. Additionally, there are at most $2 \cdot \frac{n}{2} = n$ degree 2 Steiner points in T that were incident to one of the at most $\frac{n}{2}$ arcs of N removed to form T. So there are at most $2n - 2$ Steiner points in T and thus there are at most $2n - 2$ Steiner points in N. □

There is no known minimum (A, B)-network demonstrating that this bound is tight, and so either finding such a network or improving the bound remains an open problem. It is possible a tighter bound can be found by considering only networks without degree-3 Steiner cycles. Known lower bounds on the maximum number of Steiner points for the general case and the case without degree-3 Steiner cycles are n and $n - 1$ respectively, with networks demonstrating this in Figs. 9a and 18 (right) respectively.

8 Alfaro's Conjecture

In this final section we resolve (in the negative) an open conjecture of Alfaro given in [13], relating to (A, B)-networks where $A = B$ (i.e. every terminal is both a source and a sink). We call such networks (A, A)-*networks*.

Conjecture 8.1 *(Alfaro) A minimum (A, A)-network does not have Steiner points and is a union of cycles.*

We show that a solution to the undirected 2-connected Steiner network problem in the plane can be used to obtain a minimum (A, A)-network by adding suitable directions to the edges. Given that Steiner points may occur in these networks [16], we have a counterexample to the conjecture. Such a network is given in Fig. 21. In this section we assume all networks have at least three terminals (i.e., we assume that $|A| > 2$). We also relax the assumption present since Sect. 2 that all minimum (A, B)-networks are simply embedded.

We note that in the undirected 2-connected Steiner network problem, if an optimal network S_A is 2-edge-connected it is necessarily 2-node-connected [16]. We thus refer to S_A simply as 2-connected.

Lemma 8.2 *There exists a minimum (A, A)-network with no double arcs.*

Proof Let N_A denote a minimum (A, A)-network containing a double arc. Without loss of generality we can assume that N_A is simply embedded. Then by Theorem 6.11 there is a sequence of distinct colinear nodes $u_0, u_1, \ldots, u_k$ such that for all $u_i, u_j \in V_{N_A}$, both $(u_i, u_j) \in A_{N_A}$ and $(u_j, u_i) \in A_{N_A}$ if and only if $0 \leq i \leq k - 1$ and $j = i + 1$. We also have by the theorem that nodes other than $u_0, u_1, \ldots, u_k$ do not lie on cycles. By the (A, A)-connectivity of N_A every terminal in A must lie on a cycle, and since $|A| > 2$, we have that $k > 2$. We remove the arcs (u_i, u_{i+1}) for $0 \leq i \leq n - 1$ and add an arc (u_0, u_k) to obtain a new network N'_A. Since $u_0, \ldots, u_k$ are colinear, $|N'_A| = |N_A|$. We have a dicycle around nodes $u_0, \ldots, u_k$ that is the traversal of the dipath from u_k to u_0, and the arc (u_0, u_k), so (A, A)-connectivity is maintained. There is no double arc in N'_A, as $k > 2$. Hence, N'_A is a minimum (A, A)-network with no double arcs. □

For the remainder of this section, we let N_A denote a minimum (A, A)-network containing no double arcs. By Lemma 8.2, such a network exists.

Lemma 8.3 *In the underlying undirected graph G of N_A, there are two edge disjoint paths between any two distinct nodes in A.*

Proof Suppose, by contradiction, that a pair of distinct nodes in A do not have two edge disjoint paths between them in G. Then by Menger's Theorem [7, Chapter 3] there is an edge that is a bridge in G. Thus there is an arc a that is a bridge in N_A. Let K_1 and K_2 denote the two connected components obtained by removing a from N_A. There must be at least one terminal in both K_1 and K_2, as otherwise a could be removed to shorten N_A without disrupting (A, A)-connectivity. By Lemma 6.4, at most one of K_1 and K_2 contains both a source and a sink. But every terminal is both a source and a sink, so we have a contradiction. □

Proposition 8.4 *Let S_A denote a minimum 2-connected Steiner network. Then, there exists an orientation N' of S_A such that N' is an (A, A)-network of length $|S_A|$.*

Proof We first show that adding appropriate directions to edges in S_A gives an (A, A)-network. Let N_0 denote the directed network obtained from S_A by assigning every edge an arbitrary direction. Recall Definition 4.1 defining the set $X(N)$. If $X(N_0) = A \times A$, we are done. Otherwise, we can find $u, v \in A$ such that $(u, v) \notin X(N_0)$. Since S_A is

Fig. 21 A minimum
2-connected Steiner network
containing two Steiner points,
from [16]
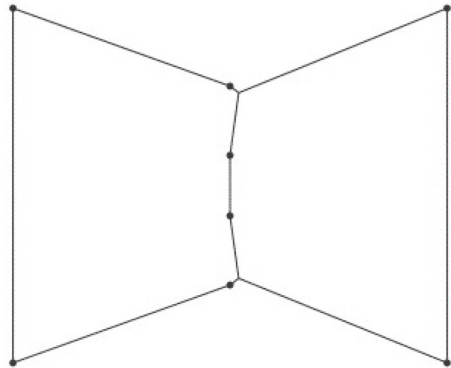
2-connected, in S_A there are two node disjoint paths from u to v, P_1 and P_2. So in N_0 we have a cycle C passing through u and v using the arcs corresponding to the edges of P_1 and P_2. Let N_1 be obtained from N_0 by orienting C to form a directed cycle. Then, in N_1 there is a dipath from u to v.

By Lemma 4.2, $X(N_0) \subseteq X(N_1)$. Since $(u, v) \in X(N_1)$ and $(u, v) \notin X(N_0)$, $|X(N_1)| > |X(N_0)|$. If $X(N_1) = A \times A$, we have an (A, A)-network. Otherwise, we repeat this process. Since at every step the size of the set of connected terminals strictly increases, it must become $A \times A$ after some finite number of iterations. So we can obtain an (A, A)-network from S_A. Furthermore, the resultant arcs correspond exactly to the original edges in the embedding, so the length of the resultant network is $|S_A|$.   □

Theorem 8.5 *Let N_A denote a minimum (A, A)-network with no double arcs. Then, the underlying undirected graph of N_A is a minimum 2-connected Steiner network on A.*

Proof Let S_A denote the underlying graph of N_A. Then, by Lemma 8.3 we know that S_A is 2-edge-connected. Suppose that S_A is not a minimum 2-connected Steiner network and let S'_A denote a minimum 2-connected Steiner network. By Proposition 8.4 we can construct an (A, A)-network N'_A of length $|S'_A|$. But then $|N'_A| = |S'_A| < |S_A| = |N_A|$, contradicting the minimality of N_A. So S_A is a minimum 2-connected Steiner network.   □

In [16], a terminal set was given such that a minimum 2-connected Steiner network in the plane contains a Steiner point. The network is given in Fig. 21. This network in conjunction with Lemma 8.2 and Theorem 8.5 shows that there is a minimum (A, A)-network containing a Steiner point, providing a counterexample to Conjecture 8.1. An example of a corresponding minimum (A, A)-network to the minimum 2-connected Steiner network in Fig. 21 would form a clockwise dicycle around the external cycle and have arcs on the chord path oriented downwards.

9 Conclusions

In this article we have proven the straightedge and compass constructibility of some minimum (A, B)-network, enabling us to present the first exact algorithmic framework for the construction of minimum (A, B)-networks in the Euclidean plane. Additionally, many structural and geometric properties of minimum (A, B)-networks have been established, including strengthening the bound on the number of Steiner points in a minimum (A, B)-network, and resolving an open conjecture of Alfaro. The problem of leveraging these properties in developing an efficient exact algorithm for the construction of minimum (A, B)-networks remains open, as does either further improving the bound on the number of Steiner points in a minimum (A, B)-network, or showing that our bound is tight.

References

1. Alfaro, Manuel: Existence of shortest directed networks in $\mathbf{R}^2$. Pac. J. Math. **167**(2), 201–214 (1995)
2. Brazil, Marcus, Graham, Ronald L., Thomas, Doreen A., Zachariasen, Martin: On the history of the Euclidean Steiner tree problem. Arch. Hist. Exact Sci. **68**(3), 327–354 (2014)
3. Brazil, Marcus, Thomas, Doreen A., Weng, Jia F.: On the complexity of the Steiner problem. J. Comb. Optim. **4**(2), 187–195 (2000)
4. Brazil, Marcus, Volz, Marcus, Zachariasen, Martin, Ras, Charl, Thomas, Doreen: Computing minimum 2-edge-connected Steiner networks in the Euclidean plane. Networks **73**(1), 89–103 (2019)
5. Brazil, Marcus, Volz, Marcus, Zachariasen, Martin, Ras, Charl, Thomas, Doreen: New pruning rules for the Steiner tree problem and 2-connected Steiner network problem. Comput. Geom. **78**, 37–49 (2019)
6. Brazil, Marcus, Martin, Zachariasen: Optimal Interconnection Trees in the Plane, pp. 151–218. Springer, London (2015)
7. Diestel, Reinhard: Graph Theory. Springer, Berlin Heidelberg (2017)
8. Feldman, Jon, Ruhl, Matthias: The directed Steiner network problem is tractable for a constant number of terminals. SIAM J. Comput. **36**(2), 543–561 (2006)
9. Feldman, Moran, Kortsarz, Guy, Nutov, Zeev: Improved approximation algorithms for directed Steiner forest. J. Comput. Syst. Sci. **78**(1), 279–292 (2012)
10. Gilbert, Edgar N., Pollak, Henry O.: Steiner minimal trees. SIAM J. Appl. Math. **1**(16), 1–29 (1968)
11. Hwang, Frank K., Richards, Dana S.: Steiner tree problems. Networks **22**(1), 55–89 (1992)
12. Juhl, Daniel, Warme, David M., Winter, Pawel, Zachariasen, Martin: The GeoSteiner software package for computing Steiner trees in the plane: an updated computational study. Math. Program. Comput. **10**(4), 487–532 (2018)

13. Maxwell, Alastair, Swanepoel, Konrad J.: Shortest directed networks in the plane. Graphs Comb. **36**(5), 1457–1475 (2020)
14. Swanepoel, Konrad: On the existence of shortest directed networks. J. Comb. Math. Comb. Comput. **33**, 97–102 (2000)
15. Winter, Pawel, Zachariasen, Martin: Euclidean Steiner minimum trees: an improved exact algorithm. Networks **30**(3), 149–166 (1997)
16. Winter, Pawel, Zachariasen, Martin: Two-connected Steiner networks: structural properties. Oper. Res. Lett. **33**(4), 395–402 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.