The Exact Complexity of ε-Dense Steiner Tree

Alper Ferudun*

September 29, 2026

Abstract

In the ε-Dense Steiner Tree problem of Karpinski and Zelikovsky, every terminal is adjacent to at least an ε-fraction of the non-terminals, and a Steiner tree with the fewest edges is sought. For every fixed $\varepsilon > 0$ the problem has a polynomial-time approximation scheme. At an Oberwolfach problem session in 2004, Hauptmann asked for hardness results and noted that it was not even known whether the exact problem is NP-hard; the question was still described as open in 2015 and in 2020. We show that for every fixed $\varepsilon \in (0, 1]$ the problem can be solved exactly in time $n^{O(\log n/\varepsilon)}$. The main step is a structural lemma: if H is any set of non-terminals that are all adjacent to terminals, and $G[S \cup H]$ has r components, then every optimal tree has at most $|H|+2r-2$ Steiner vertices adjacent to terminals. Consequently the problem is not NP-hard, even under Turing reductions, unless NP $\subseteq$ DTIME($2^{O(\log^2 n)}$). Conversely, for every fixed $\varepsilon \in (0, 1)$, a reduction from 3-SAT in the style of Megiddo and Vishkin, combined with a dense covering gadget over $\mathbb{F}_q^d$, shows that the problem has no $N^{o(\log N)}$-time algorithm unless the Exponential Time Hypothesis (ETH) fails, and that it is not in P unless FPT = W[2]. Hence, assuming ETH, exact ε-Dense Steiner Tree is neither in P nor NP-hard. Of the two halves of this statement, "not NP-hard" needs only NP $\not\subseteq$ QP, while "not in P" needs ETH or FPT $\neq$ W[2]. This is an unrefereed note.

1 Introduction

All graphs are finite, simple and undirected. An instance of Steiner Tree with unit weights is a graph $G = (V, E)$ with a set $S \subseteq V$ of *terminals*; one seeks a tree $T \subseteq G$ with $S \subseteq V(T)$ and as few edges as possible. Write $N := V \setminus S$ for the set of non-terminals. Following Karpinski and Zelikovsky [KZ98], for $\varepsilon \in (0, 1]$ the instance is ε-*dense* if $N \neq \emptyset$ and

$$|N_G(s) \cap N| \ \geq \ \varepsilon\, |N| \qquad \text{for every } s \in S. \tag{1}$$

The requirement $N \neq \emptyset$ is our addition; it is harmless, since for $N = \emptyset$ a Steiner tree exists if and only if $G[S]$ is connected. ε-Dense Steiner Tree is Steiner Tree restricted to ε-dense instances, with ε fixed. Karpinski and Zelikovsky gave a polynomial-time approximation scheme (PTAS) for it, for every fixed $\varepsilon > 0$ [KZ98].

At the Oberwolfach workshop on approximation algorithms in June 2004, Hauptmann presented approximation schemes for dense variants of Steiner problems [KKP04, pp. 1492–1495]. In the problem session [KKP04, p. 1532] he reported that efficient approximation schemes exist for ε-Dense Steiner Tree and posed, as Problem 3, the task of giving hardness results for it; he added that it was not even known whether the problem is NP-hard in the exact setting. Later work on dense Steiner problems [Hau07, Hau13, CKSV11, KLMM20] concerns mainly approximation. Cardinal, Karpinski,

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: `https://github.com/AlperTheKing`.

Schmied and Viehmann showed that Steiner Tree is APX-hard, and hence NP-hard, when the density is only polynomially small [CKSV11]; this does not apply to fixed ε (see Section 5). The online compendium of Hauptmann and Karpinski on Steiner tree problems, in its revision of April 2015, still comments in its entry on ε-Dense Steiner Tree that the problem is not known to be NP-hard in the exact setting [HK15, entry 1.6]. Karpinski, Lewandowski, Meesum and Mnich wrote in 2020 that it is open whether Steiner Tree is NP-hard on dense inputs [KLMM20, p. 2]. The question is record OWR-730-009 of the UnsolvedMath dataset [ula26].

We answer the question conditionally, in the usual complexity-theoretic sense. Here $\mathrm{QP} := \bigcup_{c\geq 1} \mathrm{DTIME}(2^{(\log n)^c})$, and ETH is the Exponential Time Hypothesis of Impagliazzo, Paturi and Zane [IP01, IPZ01]: there is $\delta > 0$ such that 3-SAT with v variables cannot be solved in time $O(2^{\delta v})$. Logarithms log are to base 2 and ln is the natural logarithm.

Theorem 1.1 (upper bound). *For every $\varepsilon \in (0, 1]$ there is an algorithm that, given an ε-dense instance with n vertices and $k \geq 2$ terminals, either finds an optimal Steiner tree or reports that none exists, in time*

$$(n+1)^c\, 3^c\, n^{O(1)}, \qquad c \leq 2\lfloor \ln k/\varepsilon\rfloor + 2\lfloor 1/\varepsilon\rfloor - 1.$$

In particular the running time is $n^{O(\log n/\varepsilon)}$.

Theorem 1.2 (lower bound). *Let $\varepsilon \in (0, 1)$ be fixed.*

*(a) Unless ETH fails, no algorithm decides in time $N^{o(\log N)}$ whether an ε-dense instance with N vertices has a Steiner tree with at most a given number of edges. In particular, if ETH holds then ε-*Dense Steiner Tree *is not solvable in polynomial time.*

*(b) ε-*Dense Steiner Tree*, parameterized by the number of Steiner vertices of the tree, is* W[2]*-hard. In particular, if ε-*Dense Steiner Tree *is solvable in polynomial time, then* FPT = W[2].

Corollary 1.3. *The following consequences hold.*

(i) If NP $\not\subseteq$ DTIME$(2^{O(\log^2 n)})$*, in particular if* NP $\not\subseteq$ QP*, then for no $\varepsilon \in (0, 1]$ is ε-*Dense Steiner Tree *NP-hard, even under polynomial-time Turing reductions.*

*(ii) If ETH holds, then for every $\varepsilon \in (0, 1)$ the problem ε-*Dense Steiner Tree *is neither solvable in polynomial time nor NP-hard.*

*(iii) For fixed $\varepsilon \in (0, 1)$, ε-*Dense Steiner Tree *has no fully polynomial-time approximation scheme unless it is solvable in polynomial time; so there is none if ETH holds or if* FPT $\neq$ W[2].

Part (i) is Corollary 3.5 below, and (ii) follows from (i) and Theorem 1.2(a), because ETH implies NP $\not\subseteq$ QP. For (iii), an optimal tree has fewer than $|V|$ edges, so a $(1 + 1/|V|)$-approximation is exact. For $\varepsilon = 1$ the problem is trivial: every non-terminal is adjacent to all terminals, so an optimal tree uses at most one Steiner vertex. An unconditional version of (ii) cannot be expected, since it would imply P $\neq$ NP. Indeed, if P = NP, then every problem in P other than the empty and the full language is NP-hard. Under ETH the exponent is right: Theorem 1.1 gives $N^{O(\log N)}$ and Theorem 1.2(a) excludes $N^{o(\log N)}$, for each fixed $\varepsilon \in (0, 1)$.

Techniques and related work. For dense Set Cover, where every element lies in at least an ε-fraction of the sets, the greedy algorithm finds a cover of size $O(\log n/\varepsilon)$, so the problem can be solved exactly by enumeration in time $m^{O(\log n/\varepsilon)}$ and is not NP-complete unless NP has quasi-polynomial algorithms; see [KZ98] and, for subdense instances, [CKSV11]. The same greedy argument shows that $O(\log n/\varepsilon)$ non-terminals dominate all terminals of a dense Steiner instance. This alone does not help, because an optimal tree may contain arbitrarily long paths of Steiner vertices that are adjacent to no terminal. Our Lemma 3.3 shows that the Steiner vertices of an optimal tree that *are* adjacent to terminals are few. One can then guess them and find the remaining part with the Dreyfus–Wagner algorithm [DW71] on $O(\log n/\varepsilon)$ contracted terminals. Lemma 3.3 does not use density at all; density enters only through the choice of the set H.

The lower bound uses a classical idea. Megiddo and Vishkin [MV88] split the variables of a formula into about $\sqrt{v}$ blocks and encode each block by $2^{O(\sqrt{v})}$ assignments, to show that a polynomial-time algorithm for minimum dominating set in tournaments would solve SAT in time $2^{O(\sqrt{v})}$. Papadimitriou and Yannakakis [PY96] placed problems of this kind in their class LOGSNP of problems with limited nondeterminism. We use the same block encoding (Lemma 4.4), a product with a dense covering gadget built from linear functionals on $\mathbb{F}_q^d$ (Lemmas 4.1 and 4.2), and a root vertex joined to all set vertices (Lemma 4.3). A root of this kind also appears in the reduction of [KLMM20] from Set Cover to dense Group Steiner Tree. As a by-product one obtains the same ETH lower bound for ε-dense Set Cover, for every $\varepsilon < 1$ (Remark 4.5). For tournaments a bound of this form is implicit in [MV88], and for dense Set Cover it may well be folklore; a related source that we did not consult is the paper of Bar-Yehuda and Kehat on dense Set Cover [BYK04]. We did not find the bound stated in the sources we checked, and we claim no novelty for the lower-bound technique.

2 Preliminaries

For $X \subseteq V$, $G[X]$ is the induced subgraph and $N_G(v)$ is the neighbourhood of v. A graph is *disconnected* if it has at least two components. A vertex v of a connected graph F is a *cut vertex* if $F - v$ is disconnected. For an instance (G, S) let $N = V \setminus S$ and let

$$D := \{v \in N : \; N_G(v) \cap S \neq \emptyset\}$$

be the set of non-terminals adjacent to at least one terminal. A set $Y \subseteq N$ is *feasible* if $G[S \cup Y]$ is connected, and

$$\mathrm{opt}(G, S) := \min\{|Y| : \; Y \subseteq N \text{ feasible}\}.$$

If $|S| \leq 1$ the problem is trivial, and if two terminals lie in different components of G no Steiner tree exists; both cases are recognised in linear time. From now on $|S| \geq 2$ and all terminals lie in one component of G, so that feasible sets exist.

Observation 2.1. *The minimum number of edges of a tree $T \subseteq G$ with $S \subseteq V(T)$ equals $|S| - 1 + \mathrm{opt}(G, S)$. From a feasible set Y one obtains such a tree with $|S| + |Y| - 1$ edges in linear time.*

Proof. If T is such a tree, then $Y := V(T) \setminus S \subseteq N$ is feasible, because $T \subseteq G[S \cup Y]$ is connected, and T has $|S| + |Y| - 1$ edges. Conversely, a spanning tree of the connected graph $G[S \cup Y]$ has $|S| + |Y| - 1$ edges. □

3 A quasi-polynomial exact algorithm

Lemma 3.1 (hubs)**.** *Let (G, S) be ε-dense with $k = |S| \geq 1$. In polynomial time one can compute a set $H' \subseteq D$ such that*

(i) every terminal has a neighbour in H';

(ii) $|H'| \leq 2\lfloor \ln k/\varepsilon \rfloor + 1$;

(iii) $G[S \cup H']$ *has at most* $\lfloor 1/\varepsilon \rfloor$ *components.*

Proof. Greedy step. Start with $H = \emptyset$. While the set U of terminals without a neighbour in H is non-empty, add to H a vertex $v \in N$ maximising $|N_G(v) \cap U|$. By (1),

$$\sum_{v \in N} |N_G(v) \cap U| \;=\; \sum_{s \in U} |N_G(s) \cap N| \;\geq\; \varepsilon\,|N|\,|U|,$$

so the chosen vertex is adjacent to at least $\varepsilon|U| > 0$ vertices of U. In particular it lies in D, and each step multiplies $|U|$ by at most $1 - \varepsilon$. After j steps $|U| \leq (1-\varepsilon)^j k \leq e^{-\varepsilon j} k$, which is less than 1 for $j = \lfloor \ln k/\varepsilon \rfloor + 1$. Hence the loop stops with $|H| \leq \lfloor \ln k/\varepsilon \rfloor + 1$, and every terminal has a neighbour in H.

Merging step. Every component of $G[S \cup H]$ contains a vertex of H, because every terminal has a neighbour in H and every vertex of H has a neighbour in S. So $G[S \cup H]$ has at most $|H|$ components. Let $X \supseteq H$ be the current set, $X \subseteq D$, and suppose that $G[S \cup X]$ has more than $t := \lfloor 1/\varepsilon \rfloor$ components. Each component contains a terminal; choose terminals $s_0, \ldots, s_t$ in $t+1$ distinct components. Since

$$\sum_{i=0}^{t} |N_G(s_i) \cap N| \;\geq\; (t+1)\,\varepsilon\,|N| \;>\; |N|,$$

two of them, say s_i and s_j, have a common neighbour $w \in N$. Then $w \notin X$, since otherwise s_i and s_j would lie in the same component, and $w \in D$. Adding w to X joins the components of s_i and s_j and creates no new component, so the number of components drops by at least one. Repeat until at most t components remain, and let H' be the final set. At most $|H| - 1$ vertices are added, so $|H'| \leq 2|H| - 1 \leq 2\lfloor \ln k/\varepsilon \rfloor + 1$. Property (i) holds because $H \subseteq H'$. All steps take polynomial time. □

Lemma 3.2 (counting)**.** *Let F be a connected graph whose vertex set is the disjoint union of sets K and Z with $|K| = r \geq 1$, such that every vertex of Z is a cut vertex of F. Then at most $2r-2$ vertices of Z have a neighbour in K. If $r = 1$ then $Z = \emptyset$.*

Proof. Let B be the set of vertices of Z with a neighbour in K, and for $z \in B$ choose a neighbour $\kappa_z \in K$. Let $M := \{z\kappa_z :\ z \in B\}$. In the graph $(V(F), M)$ every vertex of B has degree 1 and every edge has an endpoint in B, so M contains no cycle. Since F is connected, M extends to a spanning tree T of F.

Let $z \in Z$. The forest $T - z$ is a spanning subgraph of the disconnected graph $F - z$, so $T - z$ is disconnected. Hence $\deg_T(z) \geq 2$: if $\deg_T(z) = 1$ then $T - z$ is a tree, and if $\deg_T(z) = 0$ then F consists of z alone and $F - z$ is empty. In the tree T,

$$\sum_{x \in V(T)} \big(\deg_T(x) - 2\big) = 2(|V(T)| - 1) - 2|V(T)| = -2.$$

The terms with $x \in Z$ are non-negative, so $\sum_{\kappa \in K} \deg_T(\kappa) \leq 2r - 2$. Each edge of M has exactly one endpoint in K, and distinct vertices $z \in B$ give distinct edges, so $|B| = |M| \leq \sum_{\kappa \in K} \deg_T(\kappa) \leq 2r-2$. If $r = 1$, the unique vertex of K has degree 0 in the spanning tree T, so F has only one vertex and $Z = \emptyset$. □

The bound is attained: in a path whose vertices, in order, are $\kappa_1, z, z', \kappa_2, z'', z''', \kappa_3, \ldots, \kappa_r$, all $2r - 2$ vertices of Z are cut vertices adjacent to K.

Lemma 3.3 (core lemma)**.** *Let (G, S) be any instance with $|S| \geq 2$ that has a feasible set (density is not assumed). Let $H \subseteq D$, and let r be the number of components of $G[S \cup H]$. Then every optimal feasible set Y^* satisfies*

$$|Y^* \cap D| \ \leq \ |H| + 2r - 2.$$

Moreover, $G[S \cup (Y^* \cap D)]$ has at most $\max\{1, |Y^* \cap D|\}$ *components.*

Proof. Put $W := S \cup H$. The graph $G[W \cup Y^*]$ is connected, because $G[S \cup Y^*]$ is connected and every vertex of H has a neighbour in S. Choose $Z \subseteq Y^* \setminus H$ inclusion-minimal such that $G[W \cup Z]$ is connected; this is possible since $Z = Y^* \setminus H$ qualifies. The set $H \cup Z$ is feasible and $H \cap Z = \emptyset$, so by optimality $|Y^*| \leq |H| + |Z|$. As $Z \subseteq Y^*$, this gives

$$|Y^* \setminus Z| \leq |H|. \tag{2}$$

Let $W_1, \dots, W_r$ be the components of $G[W]$. Define a graph F on the vertex set $K \cup Z$, where $K = \{\kappa_1, \dots, \kappa_r\}$ is a set of new vertices: two vertices of Z are adjacent in F if they are adjacent in G, and $z \in Z$ is adjacent to κ_i if z has a G-neighbour in W_i. Thus F arises from $G[W \cup Z]$ by contracting each W_i to the vertex κ_i. Each W_i is connected and there are no edges between distinct W_i, so for every $Z' \subseteq Z$ the graph $G[W \cup Z']$ is connected if and only if $F[K \cup Z']$ is connected. Taking $Z' = Z$ shows that F is connected. For $z \in Z$, minimality of Z says that $G[W \cup (Z \setminus \{z\})]$ is disconnected, so $F - z$ is disconnected and z is a cut vertex of F. By Lemma 3.2, the set

$$B := \{z \in Z : \ N_G(z) \cap W \neq \emptyset\}$$

of vertices of Z with an F-neighbour in K has $|B| \leq 2r - 2$.

Let $y \in Y^* \cap D$. If $y \in Z$, then y has a neighbour in $S \subseteq W$, so $y \in B$. Hence $Y^* \cap D \subseteq (Y^* \setminus Z) \cup B$, and (2) gives $|Y^* \cap D| \leq |H| + 2r - 2$.

For the last claim put $C := Y^* \cap D$ and $P := Y^* \setminus C$; no vertex of P has a neighbour in S. Suppose that $G[S \cup C]$ has at least two components, and let Q be one of them. As $G[S \cup Y^*]$ is connected, some edge of $G[S \cup Y^*]$ joins $V(Q)$ to $(S \cup Y^*) \setminus V(Q)$. Its endpoint outside Q is not in $S \cup C$, because Q is a component of $G[S \cup C]$, so it lies in P. Its endpoint in Q is therefore not a terminal, so it lies in C. Thus every component of $G[S \cup C]$ contains a vertex of C, and there are at most $|C|$ components. □

Theorem 3.4. *Theorem 1.1 holds.*

Proof. Let (G, S) be ε-dense with n vertices and $k \geq 2$ terminals, all in one component of G (see Section 2). If $G[S]$ is connected, then opt $= 0$. Otherwise the algorithm proceeds as follows.

1. Compute H' by Lemma 3.1, the number r of components of $G[S \cup H']$, and $c := |H'| + 2r - 2$.

2. For every $C \subseteq D$ with $|C| \leq c$: let $Q_1, \dots, Q_m$ be the components of $G[S \cup C]$. If $m > \max\{1, |C|\}$, discard C. Otherwise compute

$$\mathrm{val}(C) := \min\{|Z'| : \ Z' \subseteq N \setminus C, \ G[S \cup C \cup Z'] \text{ connected}\}$$

together with a minimising Z'. (The minimum is finite: $C \subseteq D$ lies in the component of G that contains S, and Z' may consist of all other non-terminals of that component.) This is done as follows. Let G_C be the graph on $\{q_1, \dots, q_m\} \cup (N \setminus C)$ in which $G_C[N \setminus C] = G[N \setminus C]$, the vertices q_i are pairwise non-adjacent, and q_i is adjacent to $x \in N \setminus C$ if x has a G-neighbour in Q_i. Since each Q_i is connected and distinct Q_i are non-adjacent, $G[S \cup C \cup Z']$ is connected if and only if $G_C[\{q_1, \dots, q_m\} \cup Z']$ is connected. By Observation 2.1 applied to G_C with terminal set $\{q_1, \dots, q_m\}$, val(C) equals the minimum number of edges of a Steiner tree for $\{q_1, \dots, q_m\}$ in G_C, minus $m - 1$. This Steiner tree is computed by the Dreyfus–Wagner algorithm [DW71] in time $3^m n^{O(1)}$.

3. Output $Y := C \cup Z'$ for a candidate C minimising $|C| + \mathrm{val}(C)$, and a spanning tree of $G[S \cup Y]$.

Correctness. Every output set $C \cup Z'$ is feasible, so the output has at least opt Steiner vertices. Let Y^* be an optimal feasible set and $C^* := Y^* \cap D$. By Lemma 3.3 with $H = H'$, which is allowed because $H' \subseteq D$, we have $|C^*| \le c$ and $G[S \cup C^*]$ has at most $\max\{1, |C^*|\}$ components. So C^* is examined, and $Z' = Y^* \setminus C^*$ is admissible in the definition of $\mathrm{val}(C^*)$. Hence $|C^*| + \mathrm{val}(C^*) \le |Y^*| = \mathrm{opt}$, and the output is optimal by Observation 2.1.

Running time. The number of sets C is at most $\sum_{j \le c} \binom{|D|}{j} \le (n+1)^c$. For each of them $m \le \max\{1, c\}$, so the work is $3^c n^{O(1)}$. By Lemma 3.1, $c \le 2\lfloor \ln k/\varepsilon \rfloor + 1 + 2\lfloor 1/\varepsilon \rfloor - 2$. For fixed ε and $n \ge 3$ this is at most $(2 \ln n + 2)/\varepsilon = O(\log n/\varepsilon)$, and the total time is $n^{O(c)} = n^{O(\log n/\varepsilon)}$. ☐

Corollary 3.5. *If for some $\varepsilon \in (0, 1]$ the decision version of ε-*Dense Steiner Tree *is NP-hard under polynomial-time Turing reductions, then* NP $\subseteq$ DTIME$(2^{O(\log^2 n)})$*. In that case ETH fails.*

Proof. Compose a polynomial-time Turing reduction from SAT with the algorithm of Theorem 3.4. On inputs of length L, the reduction asks $L^{O(1)}$ queries of size $L^{O(1)}$, and each is answered in time $(L^{O(1)})^{O(\log L)} = 2^{O(\log^2 L)}$. A 3-SAT instance with v variables has, after removing repeated clauses, length polynomial in v, so it would be solved in time $2^{O(\log^2 v)} = 2^{o(v)}$, contrary to ETH. ☐

Remark 3.6. The proof of Theorem 3.4 does not need ε to be constant. If ε is part of the input, the running time is $n^{O(\log n/\varepsilon)}$, which is quasi-polynomial as long as $\varepsilon \ge 1/(\log n)^{O(1)}$. This matches the regime of the approximation results for subdense instances in [CKSV11].

4 The lower bound

A *set system* $(U, \mathcal{F})$ consists of a finite set U and a finite indexed family $\mathcal{F} = (F_i)_{i \in I}$ of subsets of U; repetitions are allowed. Its covering number $\tau(U, \mathcal{F})$ is the least $|J|$ with $J \subseteq I$ and $\bigcup_{i \in J} F_i = U$ (and ∞ if there is none). Repeating members of $\mathcal{F}$ does not change τ.

Lemma 4.1 (dense gadget)**.** *Let q be a prime and $d \ge 1$. Put $Q := \mathbb{F}_q^d \setminus \{0\}$ and, for $a \in Q$, $A_a := \{x \in Q :\ a \cdot x \ne 0\}$, where $a \cdot x = \sum_i a_i x_i$. Let $\mathcal{B}_{q,d} := (A_a)_{a \in Q}$. Then:*

(i) every $x \in Q$ lies in exactly $q^d - q^{d-1}$ of the $q^d - 1$ sets A_a, and $q^d - q^{d-1} \ge (1 - 1/q)(q^d - 1)$;

(ii) $\tau(Q, \mathcal{B}_{q,d}) = d$.

Proof. (i) For $x \ne 0$ the map $a \mapsto a \cdot x$ is a non-zero linear functional on $\mathbb{F}_q^d$, so its kernel has q^{d-1} elements. Hence exactly $q^d - q^{d-1}$ vectors a satisfy $a \cdot x \ne 0$, and all of them lie in Q. Moreover $(1 - 1/q)(q^d - 1) = q^d - q^{d-1} - 1 + 1/q \le q^d - q^{d-1}$.

(ii) A point $x \in Q$ lies in none of $A_{a_1}, \dots, A_{a_t}$ exactly when $a_i \cdot x = 0$ for all i. If $t < d$, this homogeneous system has a non-zero solution, so t sets do not cover Q. The d sets A_{e_i} for the unit vectors e_i cover Q, since every $x \ne 0$ has a non-zero coordinate. ☐

Lemma 4.2 (product)**.** *Let $(U, \mathcal{F})$ and $(Q, \mathcal{B})$ be set systems with $U, Q \ne \emptyset$, and let $\mathcal{F} \otimes \mathcal{B}$ be the family on $U \times Q$ consisting of the sets $F \times Q$ for F in $\mathcal{F}$ and $U \times A$ for A in $\mathcal{B}$. Then $\tau(U \times Q, \mathcal{F} \otimes \mathcal{B}) = \min\{\tau(U, \mathcal{F}), \tau(Q, \mathcal{B})\}$.*

Proof. If some members $F_1, \dots, F_t$ of $\mathcal{F}$ cover U, then the sets $F_j \times Q$ cover $U \times Q$; similarly for $\mathcal{B}$. So "$\le$" holds. Conversely, take a subfamily of $\mathcal{F} \otimes \mathcal{B}$, consisting of sets $F \times Q$ for F in a subfamily $\mathcal{F}'$ of $\mathcal{F}$ and sets $U \times A$ for A in a subfamily $\mathcal{B}'$ of $\mathcal{B}$. If $\mathcal{F}'$ does not cover U and $\mathcal{B}'$ does not cover Q, pick $u \in U \setminus \bigcup \mathcal{F}'$ and $x \in Q \setminus \bigcup \mathcal{B}'$; then (u, x) is not covered. So every cover satisfies $|\mathcal{F}'| \ge \tau(U, \mathcal{F})$ or $|\mathcal{B}'| \ge \tau(Q, \mathcal{B})$. ☐

Lemma 4.3 (dense Steiner instances). *Let $\varepsilon \in (0,1)$ and let q be a prime with $q > 1/(1-\varepsilon)$, so that $\rho := 1 - 1/q > \varepsilon$. Let $(U, \mathcal{F})$ be a set system with $U \neq \emptyset$ and $m := |\mathcal{F}| \geq 1$, and let $d \geq 1$. Put*

$$R := \left\lceil \frac{\varepsilon\, m}{(\rho - \varepsilon)(q^d - 1)} \right\rceil \quad (\geq 1).$$

Then one can construct, in time polynomial in its size, an ε-dense instance (G, S) with $|S| = 1 + |U|(q^d - 1)$ terminals and

$$|V(G)| = 1 + |U|(q^d - 1) + m + R(q^d - 1) \;\leq\; (|U| + 1)\, q^d + \frac{\rho\, m}{\rho - \varepsilon} + 1$$

vertices, such that $\operatorname{opt}(G, S) = \min\{\tau(U, \mathcal{F}), d\}$.

Proof. Let Q and A_a be as in Lemma 4.1. The non-terminals are a vertex f_F for each member F of $\mathcal{F}$, and R vertices $g_{a,1}, \ldots, g_{a,R}$ for each $a \in Q$. The terminals are a root o and a vertex $t_{u,x}$ for each $(u, x) \in U \times Q$. The root is adjacent to all non-terminals; $t_{u,x}$ is adjacent to f_F if $u \in F$, and to $g_{a,j}$ (for all j) if $x \in A_a$. There are no other edges. The size bound follows from $R(q^d - 1) \leq \varepsilon m/(\rho - \varepsilon) + q^d - 1$.

Density. Here $|N| = m + R(q^d - 1)$. The root sees all of N. By Lemma 4.1(i), $t_{u,x}$ sees at least $R(q^d - q^{d-1}) \geq \rho R(q^d - 1)$ non-terminals, and $\rho R(q^d - 1) \geq \varepsilon(m + R(q^d - 1))$ is equivalent to $(\rho - \varepsilon) R(q^d - 1) \geq \varepsilon m$, which holds by the choice of R.

Optimum. Associate with f_F the set $F \times Q$ and with $g_{a,j}$ the set $U \times A_a$. Then $t_{u,x}$ is adjacent to a non-terminal exactly when (u, x) lies in the associated set. Terminals are pairwise non-adjacent and $|S| \geq 2$, so if Y is feasible, every $t_{u,x}$ has a neighbour in Y; that is, the sets associated with Y cover $U \times Q$. Conversely, if the sets associated with Y cover $U \times Q$, then $Y \neq \emptyset$, every vertex of Y is adjacent to o and every $t_{u,x}$ is adjacent to a vertex of Y, so $G[S \cup Y]$ is connected. Hence $\operatorname{opt}(G, S)$ is the covering number of $\mathcal{F} \otimes \mathcal{B}'$, where $\mathcal{B}'$ is $\mathcal{B}_{q,d}$ with every member repeated R times. By Lemma 4.2 and Lemma 4.1(ii) this equals $\min\{\tau(U, \mathcal{F}), d\}$. □

Lemma 4.4 (block encoding, after [MV88]). *Let φ be a CNF formula with $v \geq 1$ variables and M clauses, and let $1 \leq k \leq v$. One can construct, in time polynomial in M and $k 2^{\lceil v/k \rceil}$, a set system $(U, \mathcal{F})$ with $|U| = M + k$ and $|\mathcal{F}| \leq k 2^{\lceil v/k \rceil}$ such that $\tau(U, \mathcal{F}) \leq k$ if and only if φ is satisfiable.*

Proof. Split the variables into non-empty blocks $X_1, \ldots, X_k$ of size at most $\lceil v/k \rceil$. Let U consist of the clauses of φ and k new elements $g_1, \ldots, g_k$. For every i and every assignment $\alpha\colon X_i \to \{0,1\}$ let $F_{i,\alpha}$ consist of g_i and the clauses that contain a literal over X_i made true by α. If β satisfies φ, the k sets $F_{i,\beta|X_i}$ cover U. Conversely, only the sets $F_{i,\cdot}$ contain g_i, so a cover with at most k sets consists of exactly one set F_{i,α_i} for each i. The assignment that agrees with α_i on X_i for every i then satisfies every clause. □

Proof of Theorem 1.2. Fix $\varepsilon \in (0,1)$ and a prime $q > 1/(1 - \varepsilon)$.

(a) Suppose an algorithm $\mathcal{A}$ decides the problem on instances with N vertices in time $N^{f(N)}$, where $f(N)/\log N \to 0$. Let φ be a 3-CNF formula with v variables. After removing repeated clauses it has $M \leq 8v^3$ distinct clauses, since each clause is a set of at most three of the $2v$ literals. Let $k := \lceil \sqrt{v} \rceil$ and $d := k + 1$. Lemmas 4.3 and 4.4 give an ε-dense instance (G, S) with N vertices and $\operatorname{opt}(G, S) = \min\{\tau(U, \mathcal{F}), k + 1\}$. Thus $\operatorname{opt}(G, S) \leq k$, that is, G has a Steiner tree with at most $|S| - 1 + k$ edges, if and only if φ is satisfiable. Since $|\mathcal{F}| \leq k 2^{\lceil v/k \rceil} \leq (\sqrt{v} + 1) 2^{\sqrt{v} + 1}$, the bound in Lemma 4.3 gives

$$2^{\sqrt{v}} \;\leq\; q^{k+1} \;\leq\; N \;\leq\; 2^{C\sqrt{v}}$$

for all $v \geq 1$, with a constant C depending only on ε and q. The instance is built in time $2^{O(\sqrt{v})}$, and running $\mathcal{A}$ takes time

$$N^{f(N)} = 2^{(f(N)/\log N)\, \log^2 N} \leq 2^{\delta(v)\, C^2 v}, \qquad \delta(v) := \sup_{N' \geq 2^{\sqrt{v}}} \frac{f(N')}{\log N'}.$$

Since $\delta(v) \to 0$, this decides 3-SAT in time $2^{o(v)}$, which contradicts ETH. A polynomial-time algorithm runs in time $N^{O(1)}$, so it is excluded as well.

(b) SET COVER parameterized by the number k of sets is W[2]-complete [DF99, CFK$^+$15]. Given $(U, \mathcal{F}, k)$ with $U \neq \emptyset$ and $\mathcal{F}$ non-empty (other cases are trivial), Lemma 4.3 with $d = k+1$ produces, in time $q^{k+1}(|U| + |\mathcal{F}|)^{O(1)}$, an ε-dense instance with opt $\le k$ if and only if $\tau(U, \mathcal{F}) \le k$, and the new parameter is again k. This is a parameterized reduction. So ε-DENSE STEINER TREE parameterized by the number of Steiner vertices is W[2]-hard. A polynomial-time algorithm for ε-DENSE STEINER TREE would make this parameterized problem fixed-parameter tractable, and then FPT = W[2]. □

Remark 4.5. The same argument gives an ETH lower bound for ε-dense SET COVER, in which the sets of the family are distinct and every element lies in at least ε times the number of sets. In Lemma 4.3 the family $\mathcal{F} \otimes \mathcal{B}'$ has repeated members (the R copies, and $A_{\lambda a} = A_a$ for $\lambda \in \mathbb{F}_q^*$), but its members occur only as neighbourhoods of distinct vertices, so this is harmless there. For SET COVER we remove the repetitions as follows. Let ε, q and ρ be as in Lemma 4.3, let $(U, \mathcal{F})$ be a set system with $U \neq \emptyset$ whose $m \ge 1$ members are distinct (removing repeated members does not change τ), let $k \ge 1$, and let $d \ge k+1$ be the least integer with $(q^d - 1)/(q-1) \ge \varepsilon m/(\rho - \varepsilon)$. Let $P \subseteq Q$ contain one vector from each one-dimensional subspace of $\mathbb{F}_q^d$, so that $|P| = (q^d - 1)/(q-1)$. The set A_a is the complement in Q of the kernel of $x \mapsto a \cdot x$, and $A_{\lambda a} = A_a$ for $\lambda \in \mathbb{F}_q^*$. Hence the sets A_a with $a \in P$ are pairwise distinct and are exactly the distinct members of $\mathcal{B}_{q,d}$, so by Lemma 4.1(ii) their covering number is d. Dividing the count in Lemma 4.1(i) by $q-1$, every $x \in Q$ lies in exactly $(q^d - q^{d-1})/(q-1) \ge \rho|P|$ of them.

On $U \times Q$ take the sets $F \times Q$ for $F \in \mathcal{F}$ and $U \times A_a$ for $a \in P$. These $m + |P|$ sets are pairwise distinct, because $U \neq \emptyset$ and $\emptyset \neq A_a \neq Q$ for $d \ge 2$. Every element lies in at least $\rho|P|$ of them, and $\rho|P| \ge \varepsilon(m + |P|)$ is equivalent to $(\rho - \varepsilon)|P| \ge \varepsilon m$, which holds by the choice of d. By Lemma 4.2 the covering number is $\min\{\tau(U, \mathcal{F}), d\}$, which is at most k if and only if $\tau(U, \mathcal{F}) \le k$. If $d > k+1$, minimality of d gives $q^{d-1} < 1 + (q-1)\varepsilon m/(\rho - \varepsilon)$, so in all cases $q^d \le \max\{q^{k+1}, q + q(q-1)\varepsilon m/(\rho - \varepsilon)\}$. Apply this to the set system of Lemma 4.4, with repeated members removed and $k = \lceil \sqrt{v} \rceil$. Then the number N of elements plus sets satisfies $2^{\sqrt{v}} \le N \le 2^{C\sqrt{v}}$ for a constant C depending only on ε and q, and the proof of Theorem 1.2(a) shows that, for every fixed $\varepsilon \in (0,1)$, ε-dense SET COVER has no $N^{o(\log N)}$-time algorithm under ETH. As said in the introduction, we do not claim novelty for this.

5 Remarks

Consistency with the approximation schemes. In the instances of Lemma 4.3 there are $|S| - 1 \ge q^d - 1$ edges that every Steiner tree must have, while opt $\le d$. Deciding whether opt $\le k$ or opt $= k+1$ therefore asks for relative accuracy of order $1/|S|$, which the PTAS of [KZ98] and the efficient PTAS reported in [KKP04, HK15] do not provide. There is no conflict with Theorem 1.2.

The parameterized view. Deciding opt $\le p$ takes time $n^{O(p)}$ by enumerating all sets of at most p non-terminals, so the problem parameterized by p is in XP; by Theorem 1.2(b) it is W[2]-hard. In the hard instances of Lemma 4.3 one has opt $\le d$, which is logarithmic in the number of vertices, so Theorem 1.2(a) already applies to instances with opt $= O(\log N)$. In general opt can be linear in n (think of long paths), but by Lemmas 3.1 and 3.3 only $O(\log n/\varepsilon)$ of the Steiner vertices of an optimal tree are adjacent to terminals.

Open questions. The dependence on ε is not settled. The upper bound has exponent $O(\log n/\varepsilon)$, while Theorem 1.2(a) only excludes exponents $o(\log N)$, and its construction does not become harder as $\varepsilon \to 0$ (for $\varepsilon < 1/2$ one may take $q = 2$). For polynomially small density the exact problem is known to be NP-hard. Cardinal, Karpinski, Schmied and Viehmann showed that, for every fixed $\delta > 0$, Steiner Tree is APX-hard, and hence NP-hard, on graphs in which every vertex has degree at least $|V \setminus S|^{1-\delta}$ [CKSV11, Theorem 5.2 of the full version]. In their instances every terminal has at least $|N|^{1-\delta}$ non-terminal neighbours, so the exact problem is NP-hard at density $\varepsilon = |N|^{-\delta}$, and hence also at $\varepsilon = n^{-\delta}$, for every fixed $\delta > 0$. (At the extreme $\varepsilon = 1/|N|$ this is elementary: the density condition then only asks every terminal to have a non-terminal neighbour, and the construction of Lemma 4.3 without the vertices $g_{a,j}$ turns Set Cover into such instances with opt $= \tau(U, \mathcal{F})$.) Together with Remark 3.6 this gives the following picture: the problem is solvable in quasi-polynomial time for $\varepsilon \geq 1/(\log n)^{O(1)}$, and it is NP-hard for $\varepsilon = n^{-\delta}$ with fixed $\delta > 0$; the regime in between, for example $\varepsilon = 2^{-\sqrt{\log n}}$, is open. (For $\varepsilon \geq n^{-o(1)}$ the algorithm of Theorem 3.4 still runs in time $2^{n^{o(1)}}$, so NP-hardness under polynomial-time reductions would contradict ETH there.) In particular the exponent must grow with $1/\varepsilon$ in some way, unless NP $\subseteq$ QP. It would be interesting to determine the right exponent as a function of both n and ε. We also did not try to decide whether the parameterized problem of Theorem 1.2(b) lies in W[2].

Verification. The proofs above are complete by hand. The following computations, which are evidence and not part of the proofs, were also made; the scripts and their outputs are in the release package.

1. The first version of the argument used the weaker bound $7r-6$ in Lemma 3.2, proved with block–cut trees. Its scripts check, by brute force and with an exact integer program, the following: the inequalities of Lemma 3.1 and of the earlier form of Lemma 3.3, and the identity $|C| + \mathrm{val}(C) =$ opt, for all 612 optimal sets of 131 small dense instances; the same on 36 larger dense instances with up to 74 non-terminals and optimum up to 29 (long paths of Steiner vertices); the counting step on 19,035 random graphs; and Lemmas 4.1 to 4.4 and the full chain 3-SAT $\to$ Set Cover $\to$ dense Steiner Tree on small cases. The values recorded in their outputs also satisfy the sharper bounds $|B| \leq 2r - 2$ and $|Y^* \cap D| \leq |H'| + 2r - 2$ used here.

2. A separate program by the author (standard library only, under one minute) checks Lemma 3.2 on all 33,867 labelled graphs with at most 6 vertices and every admissible split $K \cup Z$ (69,757 cases, K not required to be independent), and on 48,224 random cases with 7 to 12 vertices. It checks Lemma 3.3 on 942 random instances without density for *every* admissible $H \subseteq D$ (98,670 pairs); the bound is attained in some of them. It runs the algorithm of Theorem 3.4, with its own Dreyfus–Wagner implementation, on 100 dense instances (random ones, and clusters joined by long paths) and compares with brute force. It also re-checks Lemma 4.1 for nine pairs (q, d), Lemma 4.3 on 35 instances with $\varepsilon \in \{0.3, 0.6, 0.75\}$ and the full chain on 24 formulas.

3. An independent referee checked every proof of the first version and found no gap; the short proof of Lemma 3.2 given here was written afterwards, in response to the referee's suggestion to prove the bound $2r - 2$. The referee's separate code tests both bounds: Lemma 3.2 on all graphs with at most 6 vertices (with K independent, as in the application) and on 60,000 random graphs; Lemma 3.3 against all admissible H on about 33,000 random and tree-like instances, where $|H| + 2r - 2$ was never exceeded; a from-scratch version of the algorithm with brute-force completion, compared with the optimum; and the reduction, with density and optimum computed directly from the graphs.

4. A second independent referee checked every proof of the previous version of this note and found no gap; the credit to [CKSV11], the additional sources and the distinct-set version of Remark 4.5 were added in response. The referee's own program (standard library only) checks Lemma 3.2

on all labelled graphs with at most 7 vertices and every admissible split (4,033,941 cases, K arbitrary); Lemma 3.3 on all labelled graphs with at most 6 vertices, for every terminal set, every optimal Y^* and every $H \subseteq D$ (about 12.4 million pairs), and on 24,000 random sparse instances; the algorithm of Theorem 3.4 against brute force on 401 dense instances; Lemmas 3.1 and 4.1 to 4.4, the full chain on 40 formulas and the size bounds; and the gadget with one set per one-dimensional subspace used in Remark 4.5. A further short program by the author checks the construction of Remark 4.5 (distinct sets, density, and covering number $\min\{\tau(U, \mathcal{F}), d\}$) on 300 small set systems, and the chain from 3-SAT to distinct-set dense Set Cover on 40 formulas.

All runs reported 0 failures.

Scope and priority. Theorems 1.1 and 1.2 answer Hauptmann's question [KKP04, p. 1532] in the only form that is possible without separating P from NP: the exact problem is not NP-hard unless NP $\subseteq$ QP; it is not in P unless ETH fails; and it is not in P unless FPT = W[2]. The lower bound holds for each fixed $\varepsilon \in (0, 1)$; for $\varepsilon = 1$ the problem is trivial. For polynomially small density $\varepsilon = n^{-\delta}$, NP-hardness of the exact problem already follows from the APX-hardness result of Cardinal, Karpinski, Schmied and Viehmann [CKSV11, Theorem 5.2 of the full version]; see the paragraph on open questions above. Our results concern fixed ε and, for the upper bound, $\varepsilon \geq 1/(\log n)^{O(1)}$. We searched arXiv, Crossref, OpenAlex, zbMATH and the web. We found no exact complexity result for ε-Dense Steiner Tree with fixed ε. The most recent statements we found, from 2015 [HK15] and 2020 [KLMM20], describe the question as open, and OpenAlex records no citations of [KLMM20]. We did not see the full texts of [KZ98, Hau07, Hau13], of Hauptmann's dissertation [Hau04], which treats dense Steiner problems and whose zbMATH summary mentions approximation schemes for them (the full text was not accessible to us), or of the paper of Bar-Yehuda and Kehat on dense Set Cover [BYK04]. Our statements about [KZ98] follow [CKSV11, KKP04, HK15], and our description of [MV88, PY96] follows the standard secondary literature. The upper-bound method extends the known exact algorithm for dense Set Cover [KZ98, CKSV11], and the lower-bound method is that of [MV88]. The new ingredient is Lemma 3.3; the gadget of Lemma 4.1 is an elementary linear-algebra construction that we did not find used for this purpose. This negative search is not a proof of priority.

AI-assisted tools were used to search the problem corpus and the literature, to find the arguments, to write and run the verification programs, and to help prepare this manuscript. The author has reviewed the proofs, the computations and the sources and is responsible for all claims. The note has not been peer reviewed.

References

[BYK04] R. Bar-Yehuda and Z. Kehat. Approximating the dense set-cover problem. *Journal of Computer and System Sciences*, 69(4):547–561, 2004. doi:10.1016/j.jcss.2004.03.006.

[CFK+15] M. Cygan, F. V. Fomin, Ł. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms.* Springer, Cham, 2015. doi:10.1007/978-3-319-21275-3.

[CKSV11] J. Cardinal, M. Karpinski, R. Schmied, and C. Viehmann. Approximating subdense instances of covering problems. *Electronic Notes in Discrete Mathematics*, 37:297–302, 2011. Full version: arXiv:1011.0078. doi:10.1016/j.endm.2011.05.051.

[DF99] R. G. Downey and M. R. Fellows. *Parameterized Complexity.* Monographs in Computer Science. Springer, New York, 1999. doi:10.1007/978-1-4612-0515-9.

[DW71] S. E. Dreyfus and R. A. Wagner. The Steiner problem in graphs. *Networks*, 1(3):195–207, 1971. doi:10.1002/net.3230010302.

[Hau04] M. Hauptmann. *Approximation Complexity of Optimization Problems: Structural Foundations and Steiner Tree Problems*. PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät, Universität Bonn, 2004. 168 pp.; zbMATH document 5042868 (Zbl 1103.90085).

[Hau07] M. Hauptmann. On the approximability of dense Steiner tree problems. In *Theoretical Computer Science: Proceedings of the 10th Italian Conference (ICTCS 2007)*, pages 15–26. World Scientific, 2007. doi:10.1142/9789812770998_0006.

[Hau13] M. Hauptmann. On the approximability of dense Steiner problems. *Journal of Discrete Algorithms*, 21:41–51, 2013. doi:10.1016/j.jda.2013.06.005.

[HK15] M. Hauptmann and M. Karpinski. A compendium on Steiner tree problems. Online compendium, M. Hauptmann and M. Karpinski (editors), University of Bonn, 2015. Revision 288 of April 27, 2015; entry 1.6, ϵ-Dense Steiner Tree Problem; accessed September 2026. URL: http://theory.cs.uni-bonn.de/info5/steinerkompendium/netcompendium.html.

[IP01] R. Impagliazzo and R. Paturi. On the complexity of k-SAT. *Journal of Computer and System Sciences*, 62(2):367–375, 2001. doi:10.1006/jcss.2000.1727.

[IPZ01] R. Impagliazzo, R. Paturi, and F. Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001. doi:10.1006/jcss.2001.1774.

[KKP04] R. Kannan, M. Karpinski, and H. J. Prömel. Approximation algorithms for NP-hard problems. *Oberwolfach Reports*, 1(3):1457–1536, 2004. Report No. 28/2004. Problem session contribution *Steiner Tree Problems* by M. Hauptmann on pp. 1531–1532 (Problem 3 on p. 1532); abstract *PTAS for Dense Steiner Tree Problems* by M. Hauptmann on pp. 1492–1495. doi:10.4171/OWR/2004/28.

[KLMM20] M. Karpinski, M. Lewandowski, S. M. Meesum, and M. Mnich. Dense Steiner problems: Approximation algorithms and inapproximability. arXiv:2004.14102, 2020.

[KZ98] M. Karpinski and A. Zelikovsky. Approximating dense cases of covering problems. In *Network Design: Connectivity and Facilities Location (Princeton, NJ, 1997)*, volume 40 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 169–178. American Mathematical Society, Providence, RI, 1998. doi:10.1090/dimacs/040/11.

[MV88] N. Megiddo and U. Vishkin. On finding a minimum dominating set in a tournament. *Theoretical Computer Science*, 61(2–3):307–316, 1988. doi:10.1016/0304-3975(88)90131-4.

[PY96] C. H. Papadimitriou and M. Yannakakis. On limited nondeterminism and the complexity of the V-C dimension. *Journal of Computer and System Sciences*, 53(2):161–170, 1996. doi:10.1006/jcss.1996.0058.

[ula26] ulamai. UnsolvedMath. Hugging Face dataset, https://huggingface.co/datasets/ulamai/UnsolvedMath, 2026. Record OWR-730-009 (status “partially solved”); accessed September 2026.