A Logarithmic Integrality Gap for Generalizations of Quasi-Bipartite Instances of Directed Steiner Tree

Zachary Friggstad
University of Alberta, Canada

Hao Sun
University of Alberta, Canada

Abstract

In the classic Directed Steiner Tree problem (**DST**), we are given an edge-weighted directed graph $G = (V, E)$ with n nodes, a specified root node $r \in V$, and k terminals $X \subseteq V - \{r\}$. The goal is to find the cheapest $F \subseteq E$ such that r can reach any terminal using only edges in F.

Designing approximation algorithms for **DST** is quite challenging, to date the best approximation guarantee of a polynomial-time algorithm for **DST** is $O(k^\epsilon)$ for any constant $\epsilon > 0$ [Charikar et al., 1999]. For network design problems like **DST**, one often relies on natural cut-based linear programming (LP) relaxations to design approximation algorithms. In general, the integrality gap of such an LP for **DST** is known to have a polynomial integrality gap lower bound [Zosin and Khuller, 2002; Li and Laekhanukit, 2021]. So particular interest has been invested in special cases or in strengthenings of this LP.

In this work, we show the integrality gap is only $O(\log k)$ for instances of **DST** where no Steiner node has both an edge from another Steiner node and an edge to another Steiner node, i.e. the longest path using only Steiner nodes has length at most 1. This generalizes the well-studied case of quasi-bipartite **DST** where no edge has both endpoints being Steiner nodes. Our result is also optimal in the sense that the integrality gap can be as bad as poly(n) even if the longest path with only Steiner nodes has length 2.

2012 ACM Subject Classification Theory of computation → Routing and network design problems

Keywords and phrases Steiner Tree, Approximation Algorithms, Linear Programming

Digital Object Identifier 10.4230/LIPIcs.SWAT.2024.23

Funding *Zachary Friggstad*: Research supported by an NSERC Discovery Grant and Accelerator Supplement.

1 Introduction

The Directed Steiner Tree problem (**DST**) is one of the most foundational models in combinatorial optimization and network design. Given a directed graph $G = (V, E)$ with n nodes, a specified root node $r \in V$, and k terminals $X \subseteq V - \{r\}$, the goal is to buy the cheapest $F \subseteq E$ such that r can reach any terminal using only edges in F. Throughout, we say nodes in $V - (X \cup \{r\})$ are **Steiner nodes**.

Despite its central position in discrete optimization, there is a large gap in our understanding concerning its approximability. Namely, the best polynomial-time approximation is currently an $O(k^\epsilon)$-approximation for any constant $\epsilon > 0$ by Charikar et al. [4]. Grandoni, Laekhanukit, and Li show **DST** cannot be approximated within $o(\log^2 n/\log\log n)$ unless $\mathbf{NP} \subseteq \cap_{0<\delta}\mathbf{ZTIME}(2^{n^\delta})$ [12], improving on a slightly weaker lower bound than the one inherited from Group Steiner Tree [13]. These bounds differ by an order of magnitude. On the other hand, Grandoni, Laekhanukit, and Li do obtain matching $O(\log^2 k/\log\log k)$-approximation in quasi-polynomial time. Still, a polylogarithmic approximation in polynomial time remains elusive.

19th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2024).
Editor: Hans L. Bodlaender; Article No. 23; pp. 23:1–23:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1.1 Linear Programming Relaxations and Previous Work

In this paper, we consider the following natural linear programming (LP) relaxation for **DST** in which we have a variable x_e for each edge $e \in E$ modelling whether we include edge e in the solution or not.

$$\begin{array}{lrcll} \textbf{minimize}: & \sum_{e \in E} c_e \cdot x_e & & & \\ \textbf{subject to}: & x(\delta^{in}(S)) & \geq & 1 & \forall\, S \subseteq V - \{r\}, S \cap X \neq \emptyset \\ & x(\delta^{in}(v)) & \leq & 1 & \forall\, v \in V \\ & x & \geq & 0 & \end{array} \qquad (\textbf{DST-LP})$$

Here, for any $S \subseteq V$ we let $\delta^{in}(S) = \{(u,v) \in E : u \notin S, v \in S\}$ and we use the shorthand $\delta^{in}(v) := \delta^{in}(\{v\})$ for any $v \in V$. The cut constraints capture the fact that every cut separating the root from some terminal must be crossed by at least one edge in a feasible **DST** solution. In any minimal **DST** solution (i.e. a feasible $F \subseteq E$ that can not be made smaller by dropping an edge), every node will have indegree at most one since the solution is a directed tree spanning all terminals and, perhaps, some Steiner nodes. This justifies the indegree constraints. So the optimum LP solution value, denoted OPT_{LP}, is at most the cost of an optimal Steiner tree solution. We remark that (**DST-LP**) admits a simple polynomial-time separation oracle by simply checking that we can send one unit of $r - t$ flow to each terminal when edges have capacity x_e.

The integrality gap of this relaxation is well studied. First, Zosin and Khuller demonstrated the gap can be $\Omega(\sqrt{k})$ [19] in some instances. The number of vertices in their construction is exponential in the number of terminals, so the possibility of an $O(\log^c n)$ integrality gap bound was open. More recently, this was refuted by Li and Laekhanukit [15] who gave an example with integrality gap $\Omega(n^{0.0418})$. We remark that both [19] and [15] considered a different flow-based relaxation and their relaxation did not include the indegree bound for non-root nodes, but their examples are valid for (**DST-LP**).

Special Cases

Perhaps the first polylogarithimic integrality gap bound recorded for **DST** in certain settings was an $O(\log k)$ upper bound in **quasi-bipartite** instances. These are instances of **DST** such that every edge has at most one of its endpoints being a Steiner node. Another way to say this is that the subgraph induced by Steiner nodes contains no edges. Hibi and Fujito first gave an $O(\log k)$-approximation for this setting [14] and Friggstad, Könemann, and Shadravan then gave a primal-dual algorithm that demonstrated the integrality gap of (**DST-LP**) (even without the indegree constraints) is bounded by $O(\log k)$ [7]. In quasi-bipartite instances of **DST** where the underlying undirected graph excludes a fixed minor (e.g. planar graphs), (**DST-LP**) is known to have an integrality gap of $O(1)$ [8].

Chan et al. [3] generalized the $O(\log k)$ integrality gap bound to higher connectivity settings. They demonstrate an appropriate generalization of (**DST-LP**) (without the indegree constraints) for the problem of finding the cheapest $F \subseteq E$ ensuring r is at least R-edge connected to each terminal has an integrality gap bound of $O(\log k \cdot \log R)$.

Nutov [16] extended this to more settings involving more general supermodular cut requirement functions in with relaxations to the quasi-bipartite property. Namely, [16] considers a cut requirement function $f : 2^{V-\{r\}} \rightarrow \mathbb{Z}_{\geq 0}$ satisfies $f(A) + f(B) \leq f(A \cap B) + f(A \cup B)$ whenever $f(A) > 0, f(B) > 0$ and $A \cap B \cap T = \emptyset$. If one further has the property that every edge has an endpoint v such that $v \in X$ or $f(A) = 0$ for each $\{v\} \subseteq A \subseteq V - \{r\}$. In this case, [16] gives an $O(\log k \cdot \log R)$-approximation where R is the maximum value taken by f. Note, this does not capture our setting as our graphs can have edges (u, v) with both u, v being Steiner nodes yet any $A \subseteq V - \{r\}$ with $u, v \in A$ and $A \cap X \neq \emptyset$ requires an incoming edge.

Layered Graphs

An instance of **DST** is ℓ-layered if V is partitioned as $V_1 = \{r\}, V_2, V_3, \ldots, V_\ell = X$ and all edges $(u, v) \in E$ have $u \in V_i, v \in V_{i+1}$ for some $1 \leq i < \ell$. An α-approximation for **DST** in ℓ-layered graphs is known to yield an $O(\alpha \cdot \ell \cdot k^{1/\ell})$-approximation in general [18, 2]. This was the starting point for a k^ϵ-approximation by Charikar et al. [4].

The bad integrality gap examples in [19] and [15] are 5-layered instances of DST. It can easily be seen that 3-layered instances of DST (which are necessarily quasi-bipartite) have an integrality gap of $O(\log k)$ by adapting randomized Set Cover rounding techniques.

Friggstad et al. show the integrality gap of (**DST-LP**) remains $O(\log k)$ even in 4-layered instances [6]. They do this by mapping an LP solution to a natural relaxation for a related instance of Group Steiner Tree in a tree with constant height and using the known integrality gap bound for such instances [10]. Intuitively, this is possible since the first two layers of edges can only be reached in one way and each edge in the last layer is only used to connect to one terminal.

The behavior of LP relaxations for **DST** under hierarchies has also been considered. First, Rothvoss showed for ℓ-layered graphs that lifting a related flow-based LP relaxation through $O(\ell)$ layers of the Laserre hierarchy reduces the integrality gap to $O(\ell \cdot \log k)$ [17]. Later, [6] showed the result holds for a considerably weaker version of (**DST-LP**) that is valid only for layered graphs and using only the LP-based hierarchies of Lovasz-Schrijver and Sherali-Adams.

Undirected Graphs

Finally, it should be noted that in undirected graphs, the integrality gap of a related relaxation with undirected cut constraints $x(\delta(S)) \geq 1$ (and no vertex degree constraints) is well-known to be exactly 2. If one considers the bi-directed cut relaxation, i.e. the directed graph having both orientations of each undirected edge, then it is an open problem to determine if (**DST-LP**) has an integrality gap being some constant smaller than 2. It is at least known for quasi-bipartite graphs that the integrality gap of this bi-directed relaxation is better than 2 [9, 5]. Finally, a significant strengthening of the standard relaxation for general instances of undirected Steiner Tree, known as they hypergraphic relaxation, is known to have an integralty gap of $\ln(4)$ and can be efficiently solved to within any constant factor of the optimum solution cost in polynomial time [1, 11].

1.2 Our Results

We consider a generalization of **DST** in quasi-bipartite graphs and prove the following result.

▶ **Theorem 1.** *Suppose no Steiner node has both incoming and outgoing edges to other Steiner nodes. Then the integrality gap of* (**DST-LP**) *is* $O(\log k)$*.*

In other words, we consider instances where the subgraph induced by Steiner nodes may contain edges but not paths with more than one edge. Thus, this is a generalization of quasi-bipartite **DST**. This is also extends the integrality gap bound of $O(\log k)$ in 4-layered graphs [6] to a more general setting.

We emphasize that an $O(\log k)$-approximation for such graphs was already given by Hibi and Fujito [14]. The main purpose of our paper is to establish integrality gap bounds. The techniques in [14] seem unlikely to produce integrality gap bounds because they also produce $O(\log k)$-approximation for **DST** in 5-layered graphs, for which we know the integrality gap is not polylogarithmic (see Section 1.1).

Our algorithm can be seen as a common generalization of the rounding algorithm for quasi-bipartite instances from Chan et al. [3] and the analysis of Group Steiner Tree presented by Rothvoss [17]. At a high level, we round edges in phases: each phase will reduce the number of terminals we are required to connect by a constant while only paying $O(OPT_{LP})$ for the edges purchased each round.

In more detail, [3] identifies a maximal violated set around each such terminal (that excludes other required terminals) is identified and each iteration will "cover" the violated cuts in those sets. They show that no edge can be fully contained in more than one such maximal violated set around the required terminals. Unfortunately, that is not the case in our setting. Still, we can show the only edges shared between these maximal sets have at least one endpoint being a terminal, so the overlap in these sets is limited to edges between Steiner nodes. Then we use a variation of Group Steiner Tree rounding to ensure the edges e that might be used to connect to multiple nodes are only sampled with probability $O(x_e)$ in our algorithm.

2 Preliminaries

We call an edge $e = (u, v)$ a **Steiner edge** if both u and v are Steiner nodes. Call a Steiner node v a **source-Steiner** node if there is an edge (v, w) to another Steiner node w. Otherwise, call v a **sink-Steiner** node.

Recall for a subset of edges $F \subseteq E$ and a subset of nodes $S \subseteq V$ we let $\delta^{in}_F(S) = \{(u, v) \in F : u \notin S, v \in S\}$ be all edges of F entering S. Similarly, $\delta^{out}_F(S)$ are edges leaving S. If $F = E$, we may omit the subscript and simply write $\delta^{in}(S)$ and $\delta^{out}(S)$. For brevity, we also write $\delta^{in}_F(v)$ and $\delta^{out}_F(v)$ for a single node $v \in V$ to mean $\delta^{in}_F(\{v\})$ and $\delta^{out}_F(\{v\})$.

Without loss of generality, we assume there is no edge entering r (they can be deleted), no direct edge from $X \cup \{r\}$ to X (such an edge e can be subdivided with two Steiner nodes into a path of length 3 with each edge having cost $c_e/3$), and no Steiner node has no edge to any other Steiner node (such a Steiner node v can be split into two Steiner nodes v^+, v^- with a 0-cost edge from v^+ to v^-). It is straightforward to check these reductions do not change the optimal value of (**DST-LP**) and that we can map solutions between the original graph and the modified graph without increasing their costs. Again, throughout we will let OPT_{LP} denote the optimum solution value of (**DST-LP**).

2.1 Representative Terminals for Partial Solutions

Our algorithm will iteratively purchase subsets of edges over phases while making progress toward a feasible solution. So we need to understand the structure of a partial solution $F \subseteq E$ that does not necessarily connect r to each terminal. If some terminals can already reach other terminals in (V, F), we only need to focus on purchasing edges to ensure r is connected to a subset of terminals that can reach all other terminals.

For $F \subseteq E$, we consider the following **pruning** process. First, consider the strongly-connected components (SCCs) of (V, F). Since r has no incoming edges in G, then $\{r\}$ is an SCC of (V, F). Say an SCC C is a **terminal-source component** if $C \cap X \neq \emptyset$ and the only nodes in $X \cup \{r\}$ that can reach $C \cap X$ in the graph (V, F) are those already in C.

Let X_F consist of a single arbitrarily-chosen terminal in each terminal-source SCC. Note that in the graph (V, F) all terminals in X can be reached from some node in X_F but no node in X_F can be reached from any other node in X_F. To **prune** F means to iteratively remove edges from F arbitrarily as long as doing so preserves the property that every node in X can be reached from a node in $X_F \cup \{r\}$. After pruning, F looks like a directed forest where all non-singleton components have a node in $X_F \cup \{r\}$ as a root and only terminals as leaf nodes. We say F is **pruned** with respect to X_F after this process and we call X_F **representative terminals**.

▶ **Lemma 2.** *Let $F \subseteq E$ be pruned and $F' \subseteq E - F$. If $(V, F \cup F')$ contains an $r - t$ path for each $t \in X_F$, then in fact it contains an $r - t$ path for each $t' \in X$.*

Proof. Each $t \in X$ is reachable from some $t' \in X_F \cup \{r\}$ using edges in F. Since r can reach t' using edges in $F \cup F'$, it can also reach t using edges in $F \cup F'$ ◀

Additional useful properties of a pruned set of edges having roots $X_F \cup \{r\}$ are:
- Each terminal $t \in X$ can be reached from exactly one $t' \in X_F \cup \{r\}$.
- Each Steiner node u can be reached from at most one $t \in X_F \cup \{r\}$. If u can be reached this way, it is not a leaf node in its corresponding tree. If u cannot be reached from any $X_F \cup \{r\}$, it is isolated (has no incoming or outgoing edges in F).

2.2 Tracking Progress

We will find a set of edges $F' \subseteq E - F$ with cost bounded by the optimum solution value of (**DST-LP**) that, in some sense, improves overall connectivity when added to F. If we could also ensure the number of terminals not connected from r decreases by a constant factor when adding F' to F, we would be done since it would be sufficient to iterate the procedure $O(\log k)$ times.

This view too optimistic. Rather, we track progress a different way by showing $|X_F|$ decreases by a constant factor each iteration. First, we show it suffices to ensure a constant fraction of terminals in X_F can be reached by another node in $X_F \cup \{r\}$. This is essentially the same as Lemma 5 in [3], we include its proof for completeness in Appendix A.

▶ **Lemma 3.** *$0 < \alpha < 1$ and let $F' \subseteq E - F$ be such that for at least an α-fraction of $t \in X_F$, there is some other $t' \in X_F - \{t\}$ that can reach t in $(V, F \cup F')$. Then $|X_{F' \cup F}| \leq (1 - \alpha/2) \cdot |X_F|$.*

Thus, our main algorithm boils down to finding such a set F'.

▶ **Theorem 4.** *Suppose $X_F \neq \emptyset$. There is a universal constant $0 < \alpha < 1$ and a randomized algorithm with polynomial expected running time that is guaranteed to find a set $F' \subseteq E - F$ such that (a) at least an α-fraction of $t \in X_F$ are reachable from some $t' \in X_F - \{t\}$ in $(V, F \cup F')$, and (b) the cost of F' is $O(OPT_{LP})$.*

Proving Theorem 4 is the focus of Section 3.

Our final algorithm iterates the procedure from Theorem 4 and adds the resulting set F' to the current set of given edges F. Since $|X_F|$ starts at k and decreases geometrically, after $O(\log k)$ iterations the set of all edges F purchased satisfies $X_F = \emptyset$ (i.e. all terminals are reachable from r) and $cost(F) = O(\log k) \cdot OPT_{LP}$. This procedure is summarized in Algorithm 1.

Algorithm 1 **DST** Rounding.

```
Compute an optimal solution x to (DST-LP).
F ← ∅
X_F ← X ∪ {r}
while X_F ≠ {r} do
  Obtain F' ⊆ E − F using the algorithm from Theorem 4.
  F ← F ∪ F'
  Let X_F be a set of terminals, one from each source SCC in (V, F).
  Prune F with respect to X_F.
return F
```

3 The Rounding Algorithm

This section is dedicated to the proof of Theorem 4. Let x be an optimal solution to (**DST-LP**). We further assume that we cannot decrease any x_e by any positive amount.

▶ **Lemma 5.** *For each edge $e = (u, v) \in E$, $x_e \leq 1$. Additionally, if $u \neq r$ then $x_e \leq x(\delta^{in}(u))$.*

In fact, these properties would hold for any optimal solution if G had no 0-cost edges, we are just making sure 0-cost edges are well-behaved under x for our algorithm.

Proof. That $x_e \leq 1$ is obvious because all cut constraints require 1 edge, so no edge would be chosen to an extent of more than 1 in a minimal solution.

For the sake of contradiction, suppose $u \neq r$ yet $x_e > x(\delta^{in}(u))$. We claim that x_e could be decreased, contradicting minimality of x again. To see the latter, suppose otherwise, i.e. $x(\delta^{in}(S)) = 1$ for some constraint S with $e \in \delta^{in}(S)$. One easily checks

$$\begin{aligned} x(\delta^{in}(S \cup \{u\})) &= x(\delta^{in}(S)) + x(\delta^{in}(u) \cap \delta^{out}(V - S)) - x(\delta^{out}(u) \cap \delta^{in}(S)) \\ &\leq x(\delta^{in}(S)) + x(\delta^{in}(u)) - x_e \\ &< x(\delta^{in}(S)) = 1. \end{aligned}$$

This contradicts feasibility of x. ◀

Now let $F \subseteq E$ be a set of given edges (i.e. purchased in previous iterations). Our rounding procedure helps extend paths outward from nodes reachable from a node in $X_F \cup \{r\}$ toward other nodes in X_F. It does this in three phases, with the first two being very simple.

Step 1 – Forming F_1

Consider an edge $e = (u, v)$ with $u \in X \cup \{r\}$ and v being a Steiner node. Let $F_1 \subseteq E - F$ be formed by including each $e \in E - F$ independently with probability x_e. Clearly the expected cost of F_1 at most the cost of x.

Step 2 – Forming F_2

Form $F_2 \subseteq E$ as follows. For each Steiner edge $e = (u, v)$, if $\delta^{in}_{F_1}(u) \neq \emptyset$ then add f to F_2 with probability $\frac{x_e}{x(\delta^{in}(u))}$. Note the denominator cannot be 0 if we had successfully added an edge of $\delta^{in}(u)$ to F_1. So by Lemma 5, this is a valid probability. Now,

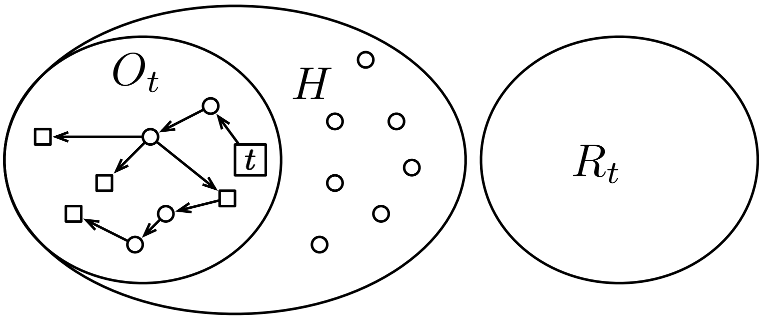

Figure 1 A depiction of the sets $O_t \cup H$ and R_t for some $t \in X_F$. Terminals are drawn as squares, Steiner nodes as circles. The edges shown are those in the pruned set F (though we do not show edges of F contained in R_t). The set R_t, which will be contracted to a single node we call r_t, consists of all nodes reachable from some other node in $X_F \cup \{r\}$ other than t. We just need to extend a path from R_t to t, the rounding algorithm we describe below will do this with constant probability.

$$\begin{aligned}
\mathbf{Pr}[e \in F_2] &= \mathbf{Pr}[e \in F_2 | \delta^{in}_{F_1}(u) \neq \emptyset] \cdot \mathbf{Pr}[\delta^{in}_{F_1}(u) \neq \emptyset] \\
&= x_e \cdot \frac{1 - \prod_{e \in \delta^{in}(u)}(1 - x_e)}{x(\delta^{in}(u))} \\
&\geq x_e \cdot \frac{1 - \exp(-x(\delta^{in}(u))}{x(\delta^{in}(u))} \\
&\geq (1 - \exp(-1)) \cdot x_e
\end{aligned}$$

The first inequality is a standard application of the arithmetic-geometric mean inequality and the bound $(1 - z/B)^B \leq \exp(-z)$ for $B \geq 1$ and $z \geq 0$. The second holds because $(1 - \exp(-1)) \cdot z \leq 1 - \exp(-z) \leq z$ holds for any $z \in [0, 1]$ and recalling the constraint $x(\delta^{in}(u)) \leq 1$ from (**DST-LP**)[1].

We also note a corresponding upper bound. The probability $\delta^{in}_{F_1}(u) \neq \emptyset$ is, by the union bound, at most $x(\delta^{in}(u))$. Using this upper bound above, we see $\mathbf{Pr}[e \in F_2] \leq x_e$.

Step 3 – Selecting the final set of edges

This step is considerably more involved, most of our new ideas are contained here. First, we discuss intuition.

Let H be all Steiner nodes v with $\delta^{in}_F(v) = \delta^{out}_F(v) = \emptyset$. For each terminal $t \in X_F$, let O_t be the set of all nodes (including t) that t can reach in (V, F). Since F is pruned, $O_t \cap O_{t'} = \emptyset$ for distinct $t, t' \in X_F$. Note that $R_t := V - (O_t \cup H)$ is the set of all nodes that can be reached by a node in $X_F - \{t\}$ using only edges in F, i.e. to reach t from $X_F - \{t\}$ it suffices to have any node in R_t reach t.

Finally, consider the graph G_t obtained by contracting R_t to a single vertex, keeping parallel edges that are created but discarding any loops. We let r_t denote this new node. Figure 1 illustrates these sets.

Now consider the following flow graph over G_t. For each edge e of G_t (i.e. an edge of G that was not contracted to a loop), install a capacity of x_e. Since $r \in R_t$, the LP constraints ensure we can send one unit of $r_t - t$ flow in G_t. We would like to sample a path from a

---

[1] This is the only point in our algorithm and analysis where we rely on this constraint.

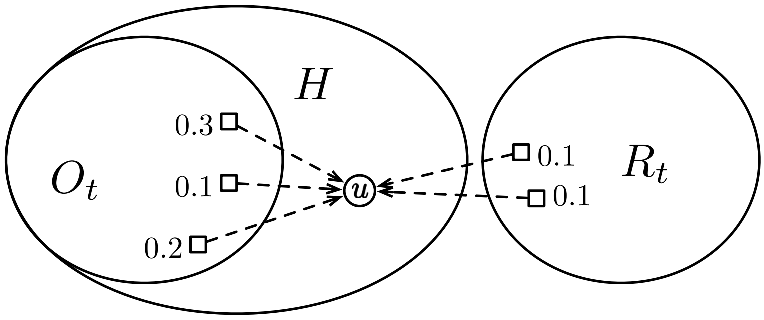

Figure 2 The Steiner node u is special for t as more than half of the value of $x(\delta^{in}(u))$ comes from nodes in O_t. Note u cannot be special for any other $t' \in X_F$ since their associated sets $O_{t'}$ are disjoint.

path decomposition of this flow, this would connect t from some other node in X_F and the expected cost of this path would be at most OPT_{LP} since no edge would be added with probability exceeding its x-value. The problem is that we cannot do this independently for different representative terminals in X_F since some edges are at risk of being considered multiple times.

We will show there is an $r_t - t$ flow of value $\geq 1/2$ that is safer to round. Intuitively, it will be that only the first two edges of any path in a path decomposition of this "safer" flow are at risk of supporting flows in G_t for too many terminals. The first two phases will have decided whether these edges will be included so we don't worry about oversampling them in this step.

Say a node $u \in H$ is **special** for terminal $t \in X_F$ if u is a source-Steiner node and the following holds:

$$\sum_{\substack{(w,u) \in \delta_G^{in}(u) \\ \text{s.t. } w \in O_t}} x_{(w,u)} > x(\delta_G^{in}(u))/2.$$

That is, u is special if more than half of the LP weight entering u comes from nodes only reachable from t. This is illustrated in Figure 2.

▷ Claim 6. Each node u is special for at most one terminal in X_F.

Proof. This is because $O_t \cap O_{t'} = \emptyset$ for distinct $t, t' \in X_F$, so at most one terminal $t \in X_F$ can have O_t claim more than half the LP weight of edges entering u. ◁

Finally, form a subgraph G'_t of G_t by including all vertices and edges except $\{(w,u) : w \in O_t$ and u is as source-steiner node that is **not** special for $t\}$. We can still push a constant amount of flow from r_t to t in G'_t, as the following shows.

▶ **Theorem 7.** *The maximum $r_t - t$ flow value in G'_t is at least $1/2$.*

Proof. For a graph G', we use notation $\delta_{G'}(S)$ to denote the set of edges of G' entering S to emphasize which graph we are discussing. Let $S \subseteq O_t \cup H$ be a subset of nodes in G'_t including t. Viewed as a subset of nodes in G, we have $x(\delta_G^{in}(S)) \geq 1$ by feasibility of the LP. Since G_t is obtained by contracting a subset of nodes lying outside of S, then $x(\delta_{G_t}(S)) \geq 1$ as well. Next we show in G'_t that this cut still has at least $1/2$ total x-weight in G'_t.

Consider any $(w,u) \in \delta^{in}(S)$. If u is not a source-Steiner or if u is special for t node then $(w,u) \in \delta^{in}_{S'}$. Otherwise, we know at least half of the weight of edges entering u comes from outside O_t, these would all be in $\delta^{in}_{G'_t}(S)$ as required. That is, $x(\delta^{in}(S)) \geq 1/2$. Since this holds for all $r_t - t$ cuts S, by the max-flow/min-cut theorem, G'_t supports at least 1/2 units of $r_t - t$ flow. ◀

Now consider any $r_t - t$ flow of value exactly 1/2 in G'_t and perform a path decomposition of this flow. That is, for various simple $r_t - t$ paths P we have a value $z_P \geq 0$ such that $\sum_P z_P = 1/2$ and $\sum_{P: e \in P} z_P \leq x_e$ for each edge e of G'_t. It is well known that such a decomposition exists with at most $|E|$ paths and can be computed in polynomial time.

Creating F_3^t

Finally we will create a set of edges F_3^t for each terminal $t \in X_F$ as follows. Consider an $r_t - t$ path P in the support of the path decomposition of G'_t. Let $E(P)$ denote the edges of G that correspond to edges of P. Let $e_1(P), e_2(P) \in E(P)$ be the first two edges of P (it may be that $|P| = 1$ in which case $e_2(P)$ is not defined). Write $e_1(P) = (v_1(P), v_2(P))$.

We consider the following random process to add some edges of $E(P)$, in doing so we also identify some **initial edges** $i(P)$ for the path P. Generally speaking, these are edges that we require to have been sampled in the formation of $F_1 \cup F_2$ in order for us to consider sampling the rest of the path P, though Case (iv) below differs slightly from this rule. Some of these cases are illustrated in Figure 3.

- **Case (i):** $v_1(P)$ **is a sink-Steiner node.**
  Set $i(P) := \emptyset$. With probability z_P, add all of $E(P)$ to F_3^t.
- **Case (ii):** $v_1(P)$ **is a source-Steiner node**
  Set $i(P) := \{e_1(P)\}$. If $e_1(P) \in F_2$, then with probability $z_P / x_{e_1(P)}$ add $E(P) - i(P)$ to F_3^t.
- **Case (iii):** $v_1(P) \in X \cup \{r\}$ **and** $v_2(P)$ **is special for** t
  Set $i(P) := \{e_1(P)\}$. If $e_1(P) \in F_1$, then with probability $z_P / x_{e_1(P)}$ add $E(P) - i(P)$ to F_3^t.
- **Case (iv):** $v_1(P) \in X \cup \{r\}$ **and** $v_2(P)$ **is not special for** t
  Then it must be that $e_2(P)$ is defined; set $i(P) := \{e_1(P), e_2(P)\}$. If $e_2(P) \in F_2$ and if **some** edge in $\delta^{in}(v_2(P)) \cap \delta^{out}(R_t)$ was added to F_1, then with probability $z_P / x_{e_2(P)}$ add $E(P) - i(P)$ to F_3^t with.

While case (ii) and (iii) are similar, there are important technical distinctions so we distinguish these cases for clarity in our analysis below. Note in all cases, if the random process adds edges of a path P to F_3^t it adds exactly the non-initial edges, i.e. $P - i(P)$.

3.1 Analysis of the Formation of the Sets F_3^t

We start by showing the probability any edge is added to a particular F_3^t is bounded by its x-value.

▶ **Lemma 8.** *For any $r_t - t$ path, the probability we added $E(P) - i(P)$ to F_3 due to processing P in its corresponding case is at most z_P. Consequently, for any $t \in X_F - \{r\}$ and any $e \in E, \mathbf{Pr}[e \in F_3^t] \leq x_e$.*

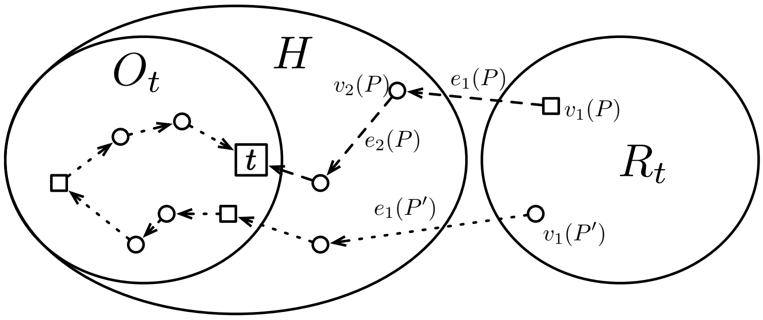

Figure 3 The graph G_t' except we have expanded node r_t to the full set R_t again. The top $r_t - t$ path (larger dashes) illustrates a path P that could either be from Case (iii) or Case (iv), depending on whether $v_2(P)$ is special for t or not. The lower path (with finer dots on the edges) illustrates a path P' from Case (ii). It might even be that some other path in the decomposition exits O_t after entering it before it eventually reaches t, but such a path could only use a Steiner edge (u, v) in H if after entering O_t if u was special for t since.

Proof. Focus on an $r_t - t$ path P and consider the corresponding case case for path P: (i) we simply added $E(P)$ with probability z_P, (ii) we added $E(P) - i(P)$ with probability $z_P/x_{e_1(P)}$ but only if $e_1(P) \in F_2$. As argued in **Step 2**, the latter happens with probability at most x_{e_1} so multiplying this against z_P/x_{e_1} finishes this case, (iii) $e_1(P)$ lies in F_1 with probability $x_{e_1(P)}$ so the total probability we added $E(P) - i(P)$ is exactly z_P.

For the final case (iv), P is sampled with probability $\frac{z_P}{x_{e_2(P)}}$ but only if the condition that includes $e_2(P) \in F_2$ is satisfied. Again, such a condition can only be satisfied with probability at most $x_{e_2}(P)$. Thus P is sampled with probability at most z_P overall.

The last statement in the lemma holds because the expected number of times an edge e is added to F_3^t is then at most $\sum_{P:e\in P} z_P \leq x_e$ because P is a path decomposition of a flow with capacity x_e on edge e. ◀

But this is not enough for a good overall cost bound, one should be concerned that an edge was added to multiple F_3^t sets for various t. The following effectively shows each edge that is a candidate to be added to some F_3^t can only support flow for at most one terminal $t \in X_F$.

▶ **Lemma 9.** *For each $e \in E$, there is at most one t such that $e \in E(P) - i(P)$ for some path P in the decomposition of the $r_t - t$ flow.*

Proof. Suppose $e = (u, v)$ has $v \in X \cup \{r\}$. The only such edges in G_t' have $v \in O_t$ since the only terminals not contracted into t_t are those in O_t. So e will only be an edge in G_t' for at most one t.

Next, suppose $u \in X \cup \{r\}$. If $u \notin O_t$ then $u \in R_t$ and $e = e_1(P)$ so we are in case (iii) or case (iv) for any path P containing e, but in either case $e \in i(P)$. Thus, we can only have $e \in E(P) - i(P)$ for the terminal t with $u \in O_t$.

Finally, suppose (u, v) is a Steiner edge. Suppose (u, v) lies on some path P in some G_t'. If $u \in R_t$ we are in case (ii) and $(u, v) \in i(P)$. If $u \notin R_t$, then either u is special for t or else the edge (w, u) prior to u is the first edge (i.e. $w \in R_t$) since we deleted all edges from O_t to u as u was not special for t. In the latter, we are in case (iv), so $(u, v) = e_2(P)$ means $(u, v) \in i(P)$. ◀

▶ **Theorem 10.** *The expected cost of $F_1 \cup F_2 \cup \bigcup_{t\in X_F - \{r\}} F_3^t$ is $O(OPT_{LP})$.*

Proof. We have already shown the expected costs of F_1 and F_2 are bounded by $O(OPT_{LP})$ since each edge is in F_1 or F_2 with probability at most x_e. We also know each e appears in any given F_3^t with probability at most x_3. Lemma 9 shows there is at most one F_3^t such that e has a nonzero probability of appearing in F_3^t, so e lies in $\bigcup_{t \in X_F - \{r\}} F_3^t$ with probability at most x_e. ◀

3.2 Success Probability

The last step is to show each terminal $t \in X_F$ can be reached from another node in $X_F - \{r\}$ with good probability. This is a bit subtle as there is shared randomness between the various $r_t - t$ paths P that reach t. Our analysis mirrors that in [17], which is providing an alternative analysis of the Group Steiner Tree rounding algorithm from [10].

We first require a general result about random variables. A proof was provided in [17] for the case $\mathbf{E}[X] = 1$. We need it in a slightly more general context so we include its proof in Appendix B for completeness.

▶ **Lemma 11.** *Let $\mu, \gamma \geq 0$ and let $X_1, X_2, \ldots, X_m$ be indicator random variables and $X = \sum_{i=1}^m X_i$ be their sum. Suppose $\mathbf{E}[X] \geq \mu$ and $\mathbf{E}[X|X_j = 1] \leq \gamma$ for any j. Then $\mathbf{Pr}[X \geq 1] \geq \mu/\gamma$.*

▶ **Theorem 12.** *There is a fixed constant $\alpha' > 0$ such that for each $t \in X_F$, with probability at least α' there is some $t' \in X_f \cup \{r\} - \{t\}$ such that t' can reach t in $(V, F \cup F')$.*

Proof. We show we added $E(P) - i(P)$ to F_3^t for at least one path P with constant probability, which suffices to prove the main result as then R_t could reach t along this path P. Consider the path decomposition and corresponding weights z_P. The subscripts in the sums on the right-hand side indicate which case the path corresponds to.

$$\frac{1}{2} = \sum_P z_P = \sum_{P:(i)} z_P + \sum_{P:(ii)} z_P + \sum_{P:(iii)} z_P + \sum_{P:(iv)} z_P$$

At least one of these sums is is at least 1/8.

Case: $\sum_{P:(i)} z_P \geq 1/8$. These paths were independently sampled with probability z_P each. The probability we did not pick one of them is then at most

$$\prod_{P:(i)} (1 - z_P) \leq \exp\left(-\sum_{P:(i)} z_P\right) \leq \exp(-1/8)$$

So at least one path was picked with probability $\geq 1 - \exp(-1/8)$.

Case: $\sum_{P:(ii)} z_P \geq 1/8$. All paths discussed here are those corresponding to case (ii) so we omit that qualifier throughout. We employ Lemma 11 where we have an indicator X_P for every path P and let $X = \sum_P X_P$. A path is added if both $e_1(P) \in F_1$ and then if P is sampled after that. This happens with probability $x_{e_1(P)} \cdot \frac{z_P}{x_{e_1(P)}} = z_P$. So $\mathbf{E}[X] \geq 1/8$.

Consider any particular path P', we want to bound $\mathbf{E}[X|X_{P'} = 1]$. We claim for any path P that

$$\mathbf{Pr}[X_P = 1 | X_{P'} = 1] = \begin{cases} 1 & \text{if } P = P' \\ z_P & \text{if } e_1(P) \neq e_1(P') \\ \frac{z_P}{x_{e_1(P')}} & \text{otherwise} \end{cases}$$

The first one is clear, the second is because $\mathbf{Pr}[X_P = 1] = \mathbf{Pr}[e_1(P) \in F_1] \cdot \frac{z_P}{x_e} = z_P$ and because the variables $X_P, X_{P'}$ are independent (since the random choice to add their initial edges F_1 were made independently). If $e_1(P) = e_1(P')$ then the only shared randomness between X_P and $X_{P'}$ was in the decision to add $e_1(P')$ to F_1. If we are given $X_{P'} = 1$, then we know $e_1(P') \in F_1$ but the choice to extend this to selecting P entirely was then made independently with probability $\frac{z_P}{x_{e_1(P')}}$.

So we have

$$\begin{aligned}
\mathbf{E}[X : X_{P'} \geq 1] &= \mathbf{Pr}[X_{P'} = 1 | X_{P'} = 1] + \sum_{P: e_1(P) \neq e_1(P')} \mathbf{Pr}[X_P = 1 | X_{P'} = 1] \\
&\quad + \sum_{P: P \neq P' \text{ and } e_1(P) = e_1(P')} \mathbf{Pr}[X_P = 1 | X_{P'} = 1] \\
&= 1 + \sum_{P: e_1(P) \neq e_1(P')} z_P + \sum_{P: P \neq P' \text{ and } e_1(P) = e_1(P')} \frac{z_P}{x_{e_1(P')}} \\
&\leq 1 + \frac{1}{2} + \frac{x_{e_1(P')}}{x_{e_1(P')}} \\
&= 5/2
\end{aligned}$$

That is, the total weight of all paths in the decomposition is at most the value of the flow, which is $1/2$. Similarly, the total weight of all paths including the edge $e_1(P')$ is at most $x_{e_1(P')}$ since the flow respects capacities.

Using Lemma 11 with $\mu = 1/8$ and $\gamma = 5/2$ shows at least one path is sampled with probability at least $1/20$.

Case: $\sum_{P:(iii)} z_P \geq 1/8$. The proof is essentially identical to the previous case and is omitted. We get the probability at least one path is sampled is at least $1/20$.

Case: $\sum_{P:(iv)} z_P \geq 1/8$. Use similar indicator variables X_P and their sum X as in case (ii), but this time for the paths of form (iv). For any such path P, we have

$$\begin{aligned}
\mathbf{Pr}[X_P = 1] &= \frac{z_P}{x_e} \cdot \mathbf{Pr}[\delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset \wedge e_2(P) \in F_2] \\
&= \frac{z_P}{x_e} \cdot \mathbf{Pr}[\delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset] \\
&\quad \cdot \mathbf{Pr}[e_2(P) \in F_2 | \delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset] \\
&= \frac{z_P}{x_e} \cdot \mathbf{Pr}[\delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset] \cdot \frac{x_e}{x(\delta^{in}(u))} \\
&= \frac{z_P}{x(\delta^{in}(u))} \cdot \mathbf{Pr}[\delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset]
\end{aligned}$$

For brevity, let $B = \delta^{in}(v) \cap \delta^{out}(R_t)$. The last probability is

$$1 - \prod_{e \in B} (1 - x_e) \geq 1 - \exp\left(-\sum_{e \in B} x_e\right) \geq 1 - \exp(-x(\delta^{in}(v))/2)$$

The final inequality is because $v_2(P)$ is not special for t. For $z \in [0, 1]$, we have[2] $1 - \exp(-z/2) \geq (1 - \exp(-1/2)) \cdot z$, so the last expression is at least $(1 - \exp(-1/2)) \cdot x(\delta^{in}(u))$ and we finally see $\mathbf{Pr}[X_P = 1] \geq (1 - \exp(-1/2)) \cdot z_P$. Thus, $\mathbf{E}[X] \geq \frac{1-\exp(-1/2)}{8}$.

[2] This holds since $1 - \exp(-z/2) = (1 - \exp(-1/2)) \cdot z$ for $z \in \{0, 1\}$ and since $1 - \exp(-z/2)$ is concave.

Finally, we upper bound $\mathbf{E}[X|X_{P'}=1]$ by a constant for any path P' considered in this case. Partition the set of paths from this case (iv) into four sets essentially based on how they interact with P' along their prefixes: $\{P'\}, \mathcal{P}_0 = \{P : v_2(P) \neq v_2(P')\}, \mathcal{P}_1 = \{P : v_2(P) = v_2(P')$ yet $e_2(P) \neq e_2(P')\}$, and $\mathcal{P}_2 = \{P : e_2(P) = e_2(P')\}$.

For $P \in \mathcal{P}_0$, simple inspection shows X_P and $X_{P'}$ are independent random variables so $\sum_{P\in\mathcal{P}_0} \mathbf{Pr}[X_P = 1|X_{P'} = 1] = \sum_{P\in\mathcal{P}_0} \mathbf{Pr}[X_P = 1] \leq \sum_{P\in\mathcal{P}_0} z_P \leq \frac{1}{2}$.

For $P \in \mathcal{P}_1$, we are given $\delta^{in}(v_2(P)) \cap \delta^{out}(R_t) \cap F_1 \neq \emptyset$ since $X_{P'} = 1$, so

$$\begin{aligned}\mathbf{Pr}[X_P = 1|X_{P'} = 1] &= \frac{z_P}{x_{e_2(P)}} \cdot \mathbf{Pr}[e_2(P) \in F_2|X_{P'} = 1] \\ &= \frac{z_P}{x_{e_2(P)}} \cdot \frac{x_{e_2(P)}}{x(\delta^{in}(v_2(P))} \\ &= \frac{z_P}{x(\delta^{in}(v_2(P))}.\end{aligned}$$

The total flow passing through $v_2(P)$ is at most its incoming edge capacity, so summing over all $P \in \mathcal{P}_1$ shows $\sum_{P\in\mathcal{P}_1} \mathbf{Pr}[X_P = 1|X_{P'} = 1] \leq 1$.

For $P \in \mathcal{P}_2$, we simply have $\mathbf{Pr}[X_P = 1|X_{P'} = 1] = \frac{z_P}{x_{e_2(P)}}$ since the condition to be met before sampling P is satisfied if we are given $X_{P'} = 1$. So $\sum_{P\in\mathcal{P}_2} \mathbf{Pr}[X_P = 1|X_{P'} = 1] = \sum_{P\in\mathcal{P}_2} \frac{z_P}{x_{e_2(P)}} \leq 1$. Thus,

$$\sum_P \mathbf{Pr}[X_P = 1|X_{P'} = 1] = 1 + \sum_{i\in\{0,1,2\}} \sum_{P\in\mathcal{P}_i} \mathbf{Pr}[X_P = 1|X_{P'} = 1] \leq 1 + \frac{1}{2} + 1 + 1 = 7/2.$$

Using Lemma 11 with $\mu = \frac{1-\exp(-1/2)}{8}$ and $\gamma = 7/2$ shows in the probability at least one path is sampled is at least some universal constant. Summarizing, no matter which case has at least 1/8 of the weight of paths we see there is a constant probability at least one path will be sampled. This completes the proof. ◀

We have shown the expected cost of the set $F' := F_1 \cup F_2 \cup \bigcup_{t\in X_F-\{r\}} F_3^t$ is at most $c \cdot OPT_{LP}$ for some universal constant c. We also showed each terminal $t \in X_F - \{r\}$ will be reachable from some other $t' \in X_t - \{t\}$ with probability at least some universal constant $\alpha' > 0$. So the expected number of terminals of this kind is at least $\alpha' \cdot |X_F|$.

Say this procedure failed if the cost of F' exceeds $\Delta \cdot c \cdot OT_{LP}$ for some constant Δ to be determined soon, or if the number of representative terminals that are now reachable from another representative is smaller than $\frac{\alpha'}{2} \cdot |X_F|$. Note we can check this condition in polynomial time.

The former happens with probability at most $1/\Delta$ by Markov's inequality. A standard variant of Markov's inequality for lower tails shows that if Y is a random variable with $\mathbf{E}[Y] \geq \alpha' \cdot M$ where M is the maximum possible value of Y, then $\mathbf{Pr}[Y < \frac{\alpha'}{2} \cdot M] \leq \frac{1-\alpha'}{1-\alpha'/2} < 1$. In our setting, we let Y be the number of representative terminals that become connected from another node in X_t after buying F', so the maximum value of Y is $|X_t|$ and the expected value is at least $\alpha \cdot |X_t|$.

Thus, by the union bound the procedure fails with probability at most $\frac{1}{\Delta} + \frac{1-\alpha}{1-\alpha/2}$. For sufficiently large constant Δ depending only on α, this is a constant less than one. That is, the procedure succeeds with constant probability. The final randomized algorithm then iterates this procedure until it does not fail, the expected number of iterations is constant. This proves Theorem 4.

References

1 Jarosław Byrka, Fabrizio Grandoni, Thomas Rothvoss, and Laura Sanità. Steiner tree approximation via iterative randomized rounding. *J. ACM*, 60(1), 2013.

2 Gruia Calinescu and Alexander Zelikovsky. The polymatroid steiner problems. *J. Combonatorial Optimization*, 33(3):281–294, 2005.

3 Chun-Hsiang Chan, Bundit Laekhanukit, Hao-Ting Wei, and Yuhao Zhang. Polylogarithmic approximation algorithm for k-connected directed steiner tree on quasi-bipartite graphs. *arXiv preprint*, 2019. `arXiv:1911.09150`.

4 Moses Charikar, Chandra Chekuri, To-Yat Cheung, Zuo Dai, Ashish Goel, Sudipto Guha, and Ming Li. Approximation algorithms for directed steiner problems. *Journal of Algorithms*, 33(1):73–91, 1999.

5 Andreas Emil Feldmann, Jochen Könemann, Neil Olver, and Laura Sanità. On the equivalence of the bidirected and hypergraphic relaxations for steiner tree. *Mathematical Programming*, 160:379–406, 2014.

6 Zachary Friggstad, Jochen Könemann, Young Kun-Ko, Anand Louis, Mohammad Shadravan, and Madhur Tulsiani. Linear programming hierarchies suffice for directed steiner tree. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 285–296. Springer, 2014.

7 Zachary Friggstad, Jochen Könemann, and Mohammad Shadravan. A Logarithmic Integrality Gap Bound for Directed Steiner Tree in Quasi-bipartite Graphs . In Rasmus Pagh, editor, *15th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2016)*, volume 53 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:11, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

8 Zachary Friggstad and Ramin Mousavi. A Constant-Factor Approximation for Quasi-Bipartite Directed Steiner Tree on Minor-Free Graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023)*, pages 13:1–13:18, 2023.

9 Isaac Fung, Konstantinos Georgiou, Jochen Könemann, and Malcolm Sharpe. Efficient algorithms for solving hypergraphic steiner tree relaxations in quasi-bipartite instances. *CoRR*, abs/1202.5049, 2012. `arXiv:1202.5049`.

10 Naveen Garg, Goran Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group steiner tree problem. *Journal of Algorithms*, 37(1):66–84, 2000.

11 Michel X. Goemans, Neil Olver, Thomas Rothvoß, and Rico Zenklusen. Matroids and integrality gaps for hypergraphic steiner tree relaxations. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, pages 1161–1176, 2012.

12 Fabrizio Grandoni, Bundit Laekhanukit, and Shi Li. O (log2 k/log log k)-approximation algorithm for directed steiner tree: a tight quasi-polynomial-time algorithm. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 253–264, 2019.

13 Eran Halperin and Robert Krauthgamer. Polylogarithmic inapproximability. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 585–594, 2003.

14 Tomoya Hibi and Toshihiro Fujito. Multi-rooted greedy approximation of directed steiner trees with applications. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 215–224. Springer, 2012.

15 Shi Li and Bundit Laekhanukit. Polynomial integrality gap of flow lp for directed steiner tree. *arXiv preprint*, 2021. `arXiv:2110.13350`.

16 Zeev Nutov. On rooted k-connectivity problems in quasi-bipartite digraphs. In *Computer Science – Theory and Applications: 16th International Computer Science Symposium*, pages 339–348. Springer-Verlag, 2021.

17 Thomas Rothvoß. Directed steiner tree and the lasserre hierarchy. *arXiv preprint*, 2011. `arXiv:1111.5473`.

18 Alexander Zelikovsky. A series of approximation algorithms for the acyclic directed steiner tree problem. *Algorithmica*, 18(1):99–110, 1997.

19 Leonid Zosin and Samir Khuller. On directed steiner trees. In *SODA*, volume 2, pages 59–63. Citeseer, 2002.

A Proof of Lemma 3

Proof. For each $t \in X_F$ that can be reached from some other $t' \in X_F \cup \{r\} - \{t\}$ in $(V, F \cup F')$, let $d(t) = t'$. If t can be reached from multiple such t', pick one arbitrarily to be $d(t)$. Finally, let F^* be all such edges $(d(t), t)$. We note that $(V, F \cup F')$ and $(V, F \cup F' \cup F^*)$ have the same SCCs because F^* provides direct connections between nodes that were already reachable in $(V, F \cup F')$.

We add the edges of F^* one at a time to (V, F) and track how the number of terminal-source SCCs decreases. Recall an SCC of (V, F) is a strongly connected component C containing a terminal that cannot be reached from any other terminal apart from those in C.

When adding $e_t = (d(t), t)$, let $S_{d(t)}$ and S_t be the SCCs containing $d(t)$ and t respectively at that time. We note S_t was a source SCC just before adding e_t because no edge entered the source component containing t before this addition.

If the number of terminal-source SCCs does not decrease after adding S_t, it must have been that S_t could already reach $S_{d(t)}$ by some path P. Let e' be the edge entering $S_{d(t)}$. Note $e' \in F^*$ since no vertex outside of $d(t)$'s SCC in (V, F) could reach $d(t)$ before (as it was a source SCC). Also note that e_t and e' are now drawn into the same SCC as S_t after e_t is added so e' will never enter another SCC again as we continue adding edges of F^*. That is, the number of iterations of adding an edge of the form e_t that do not cause the number of source SCCs to drop is at most $\alpha/2 \cdot |X_F|$, meaning the number of source SCCs in $(V, F \cup F^*)$ is at most $(1 - \alpha/2) \cdot |X_F|$. Thus, the number of terminal-source SCCs in $(V, F \cup F')$ is also bounded by $(1 - \alpha/2) \cdot |X_F|$ as required. ◀

B Proof of Lemma 11

Proof. This proof essentially just verifies the arguments in [17] generalize as required. Including the proof here also keeps our paper self-contained.

We do this in two steps. First, suppose we knew $\mathbf{E}[X|X \geq 1] \leq \gamma$. Then

$$\mu \leq \mathbf{E}[X] = \mathbf{E}[X|X = 0] \cdot \mathbf{Pr}[X = 0] + \mathbf{E}[X|X \geq 1] \cdot \mathbf{Pr}[X \geq 1] \leq \gamma \cdot \mathbf{Pr}[X \geq 1].$$

Rearranging shows $\mathbf{Pr}[X \geq 1] \geq \mu/\gamma$ which is what we wanted to show.

Now we show $\mathbf{E}[X|X \geq 1] \leq \gamma$ follows if $\mathbf{E}[X|X_j = 1] \leq \gamma$ for any j. By Jensen's inequality applied to the conditioned distribution, we have

$$\begin{aligned}
\mathbf{E}[X|X \geq 1]^2 &\leq \mathbf{E}[X^2|X \geq 1] \\
&= \sum_{(i,j)} \mathbf{Pr}[X_i = 1 \wedge X_j = 1 | X \geq 1] \\
&= \sum_{(i,j)} \mathbf{Pr}[X_j = 1 | X \geq 1 \wedge X_i = 1] \cdot \mathbf{Pr}[X_i = 1 | X \geq 1] \\
&= \sum_{(i,j)} \mathbf{Pr}[X_j = 1 | X_i = 1] \cdot \mathbf{Pr}[X_i = 1 | X \geq 1] \\
&= \sum_{i} \mathbf{Pr}[X_i = 1 | X \geq 1] \cdot \sum_{j} \mathbf{Pr}[X_j = 1 | X_i = 1] \\
&= \sum_{i} \mathbf{Pr}[X_i = 1 | X \geq 1] \cdot \mathbf{E}[X | X_i = 1] \\
&\leq \gamma \cdot \sum_{i} \mathbf{Pr}[X_i = 1 | X \geq 1] \\
&= \gamma \cdot \mathbf{E}[X | X \geq 1]
\end{aligned}$$

All sums over (i, j) are over all m^2 ordered pairs of indices. To conclude, $\mathbf{E}[X|X \geq 1]^2 \leq \gamma \cdot \mathbf{E}[X|X \geq 1]$ and $\gamma \geq 0$ can only happen if $\mathbf{E}[X|X \geq 1] \leq \gamma$. ◀