Lower Bounds on 0-Extension with Steiner Nodes

Yu Chen
EPFL, Lausanne, Switzerland

Zihan Tan
Rutgers University, Piscataway, NJ, USA

Abstract

In the 0-*Extension problem*, we are given an edge-weighted graph $G = (V, E, c)$, a set $T \subseteq V$ of its vertices called terminals, and a semi-metric D over T, and the goal is to find an assignment f of each non-terminal vertex to a terminal, minimizing the sum, over all edges $(u, v) \in E$, the product of the edge weight $c(u, v)$ and the distance $D(f(u), f(v))$ between the terminals that u, v are mapped to. Current best approximation algorithms on 0-Extension are based on rounding a linear programming relaxation called the *semi-metric LP relaxation*. The integrality gap of this LP, is upper bounded by $O(\log |T|/\log\log |T|)$ and lower bounded by $\Omega((\log |T|)^{2/3})$, has been shown to be closely related to the quality of cut and flow vertex sparsifiers.

We study a variant of the 0-Extension problem where Steiner vertices are allowed. Specifically, we focus on the integrality gap of the same semi-metric LP relaxation to this new problem. Following from previous work, this new integrality gap turns out to be closely related to the quality achievable by cut/flow vertex sparsifiers with Steiner nodes, a major open problem in graph compression. We show that the new integrality gap stays superconstant $\Omega(\log\log |T|)$ even if we allow a super-linear $O(|T| \log^{1-\varepsilon} |T|)$ number of Steiner nodes.

2012 ACM Subject Classification Theory of computation → Sparsification and spanners

Keywords and phrases Graph Algorithms, Zero Extension, Integrality Gap

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.47

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/pdf/2401.09585.pdf`

Funding *Zihan Tan*: Supported by a grant to DIMACS from the Simons Foundation (820931).

Acknowledgements We would like to thank Julia Chuzhoy for many helpful discussions. We want to thank Arnold Filtser for pointing to us some previous works on similar problems.

1 Introduction

In the 0-*Extension* problem (0-Ext), we are given an undirected edge-weighted graph $G = (V, E, c)$, a set $T \subseteq V$ of its vertices called *terminals*, and a metric D on terminals, and the goal is to find a mapping $f : V \to T$ that maps each vertex to a terminal in T, such that each terminal is mapped to itself (i.e., $f(t) = t$ for all $t \in T$), and the sum $\sum_{(u,v)\in E} c(u, v) \cdot D(f(u), f(v))$ is minimized.

The 0-Ext problem was first introduced by Karzanov [22]. It is a generalization of the *multi-way cut* problem (by setting $D(t, t') = 1$ for all pairs $t, t' \in T$) [15, 7, 17, 5, 2, 6, 4], and a special case of the *metric labeling* problem [23, 10, 3, 20, 14]. Călinescu, Karloff and Rabani [8] gave the first approximation algorithm for 0-Ext, achieving a ratio of $O(\log |T|)$, by rounding the solution of a *semi-metric LP relaxation* (LP-Metric), which is presented below.

51st International Colloquium on Automata, Languages, and Programming (ICALP 2024).
Editors: Karl Bringmann, Martin Grohe, Gabriele Puppis, and Ola Svensson;
Article No. 47; pp. 47:1–47:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$$\text{(LP-Metric)} \quad \text{minimize} \quad \sum_{(u,v)\in E} c(u,v)\cdot\delta(u,v)$$
$$s.t. \quad (V,\delta) \text{ is a semi-metric space}$$
$$\delta(t,t') = D(t,t'), \quad \forall t,t' \in T$$

Fakcharoenphol, Harrelson, Rao and Talwar [16] later gave a modified rounding algorithm on the same LP, improving the ratio to $O(\log|T|/\log\log|T|)$, which is the current best-known approximation. On the other hand, this LP was shown to have integrality gap $\Omega(\sqrt{\log|T|})$ [8], and this was recently improved to $\Omega((\log|T|)^{2/3})$ by Schwartz and Tur [31]. Another LP relaxation called *earthmover distance relaxation* (LP-EMD) was considered by Chekuri, Khanna, Naor and Zosin [10] and utilized to obtain an $O(\log|T|)$-approximation of the metric labeling problem (and therefore also the 0-Ext problem). It has been shown [21] by Karloff, Khot, Mehta and Rabani that this LP relaxation has an integrality gap $\Omega(\sqrt{\log|T|})$. Manokaran, Naor, Raghavendra and Schwartz [28] showed that the integrality gap of this LP relaxation leads to a hardness of approximation result, assuming the Unique Game Conjecture.

In addition to being an important problem on its own, the 0-Ext problem and its two LP relaxations are also closely related to the construction of *cut/flow vertex sparsifiers*, a central problem in the paradigm of graph compression. Given a graph G and a set $T \subseteq V(G)$ of terminals, a cut sparsifier of G with respect to T is a graph H with $V(H) = T$, such that for every partition (T_1, T_2) of T, the size of the minimum cut separating T_1 from T_2 in G and the size of the minimum cut separating T_1 from T_2 in H, are within some small factor q, which is also called the *quality* of the sparsifier[1]. Moitra [29] first showed that every graph with k terminals admits a cut sparsifier with quality bounded by the *integrality gap* of its LP-Metric (hence $O(\log k/\log\log k)$). Later on, Leighton and Moitra [26], and Makarychev and Makarychev [27] concurrently obtained the same results for flow sparsifiers, and then Charikar, Leighton, Li and Moitra [9] showed that the best flow sparsifiers can be computed by solving an LP similar to LP-EMD. On the lower bound side, it was shown after a line of work [26, 9, 27] that there exist graphs with k terminals whose best flow sparsifier has quality $\tilde{\Omega}(\sqrt{\log k})$.

A major open question on vertex sparsifiers is:

Q1. Can better quality sparsifiers be achieved by allowing a small number of Steiner vertices?

In other words, what if we no longer require that the sparsifier H only contain terminals, but just require that H contain all terminals and its size be bounded by some function f on the number of terminals (for example, $f(k) = 2k, k^2$ or even 2^k)? Chuzhoy [13] constructed $O(1)$-quality cut/flow sparsifiers with size dependent on the number of terminal-incident edges in G. Andoni, Gupta and Krauthgamer [1] showed the construction for $(1+\varepsilon)$-quality flow sparsifiers for quasi-bipartite graphs. For general graphs, they constructed a sketch of size $f(k,\varepsilon)$ that stores all feasible multicommodity flows up to a factor of $(1+\varepsilon)$, raising the hope for a special type of $(1+\varepsilon)$-quality flow sparsifier, called contraction-based flow sparsifiers, of size $f(k,\varepsilon)$ for general graphs, which was recently invalidated by Chen and Tan [12], who showed that contraction-based flow sparsifiers whose size are bounded by any function $f(k)$ must have quality $1+\Omega(1)$. But it is still possible for such flow sparsifiers with constant quality and finite size to exist. Prior to this work, Krauthgamer and Mosenzon [24] showed that there exist 6-terminal graphs G whose quality-1 flow sparsifiers must have an arbitrarily large size.

[1] flow sparsifiers has a slightly more technical definition, which can be found in [19, 26, 13, 1].

Given the concrete connection between the 0-Ext problem and cut/flow sparsifiers, it is natural to ask a similar question for 0-Ext:

Q2. Can better approximation of 0-Ext *be achieved by allowing a small number of Steiner vertices?*

In this paper, we formulate and study the following variant of the 0-Ext problem, which we call the *0-Extension with Steiner Nodes* problem (0EwSN). (We note that a similar variant was mentioned in [1], and we provide a comparison between them in more detail in Appendix A.) We are also given a function $f : \mathbb{Z} \to \mathbb{Z}$ with $f(k) \geq k$ for all $k \in \mathbb{Z}$, this should be the total number of Steiner vertices.

0-Extension with Steiner Nodes

In an instance of $\mathsf{0EwSN}(f)$, the input consists of an edge-weighted graph $G = (V, E, c)$, a subset $T \subseteq V$ of k vertices, that we call *terminals*, and a metric D on terminals in T, which is exactly the same as 0-Ext. A solution to the instance (G, T, D) consists of

- a partition $\mathcal{F}$ of V with $|\mathcal{F}| \leq f(k)$, such that distinct terminals of T belong to different sets in $\mathcal{F}$; we call sets in $\mathcal{F}$ *clusters*, and for each vertex $u \in V$, we denote by $F(u)$ the cluster in $\mathcal{F}$ that contains it;
- a semi-metric δ on the clusters in $\mathcal{F}$, such that for each pair $t, t' \in T$, $\delta(F(t), F(t')) = D(t, t')$.

We define the *cost* of a solution $(\mathcal{F}, \delta)$ as $\mathsf{cost}(\mathcal{F}, \delta) = \sum_{(u,v)\in E} c(u, v) \cdot \delta(F(u), F(v))$, and its *size* as $|\mathcal{F}|$. The goal is to compute a solution $(\mathcal{F}, \delta)$ with size at most $f(k)$ and minimum cost.

The difference between $\mathsf{0EwSN}(f)$ and 0-Ext is that, instead of enforcing every vertex to be mapped to a terminal, in $\mathsf{0EwSN}(f)$ we allow vertices to be mapped to $(f(k) - k)$ non-terminals (or Steiner nodes), which are the clusters in $\mathcal{F}$ that do not contain terminals. We are also allowed to manipulate the distances between these non-terminals, conditioned on not destroying the induced metric D on terminals. Clearly, when $f(k) = k$, the $\mathsf{0EwSN}(f)$ problem degenerates to the 0-Ext problem.

It is easy to see that (LP-Metric) is still an LP relaxation for $\mathsf{0EwSN}(f)$, as each solution $(\mathcal{F}, \delta)$ to 0EwSN naturally corresponds to a semi-metric δ' on V (where we can set $\delta'(u, u') = \delta(F(u), F(u'))$ for all pairs $u, u' \in V$). Denote by $\mathsf{IG}_f(k)$ the worst integrality gap for (LP-Metric) to any $\mathsf{0EwSN}(f)$ instance with at most k terminals. In fact, similar to the connection between the integrality gap of (LP-Metric) and the quality achievable by flow sparsifiers [29, 26], it was recently shown by Chen and Tan [12] that the value of $\mathsf{IG}_f(k)$ is also closely related to the quality achievable by flow sparsifiers with Steiner nodes. Specifically, for any function f, every graph G with k terminals has a quality-$\big((1 + \varepsilon) \cdot \mathsf{IG}_f(k)\big)$ flow sparsifier with size bounded by $(f(k))^{(\log k/\varepsilon)^{k^2}}$. This means that any positive answer to question Q2 (by proving that $\mathsf{IG}_f(k) = o(\log k/ \log\log k)$ for some f) also gives a positive answer to question Q1.

This makes it tempting to study the 0EwSN problem. Specifically, can we prove any better-than-$O(\log k/ \log\log k)$ upper bound for $\mathsf{IG}_f(k)$, for any function f? To the best of our knowledge, no such bound is known for any f, leaving the problem wide open. In fact, no non-trivial lower bound on $\mathsf{IG}_f(k)$ is known for even very small function like $f(k) = O(k)$.

1.1 Our Results

In this paper, we make a first step in investigating the value of $\mathsf{IG}_f(k)$, by giving a superconstant lower bound on $\mathsf{IG}_f(k)$ for near-linear functions f. Our main result is summarized in the following theorem.

▶ **Theorem 1.** *For any $0 < \varepsilon < 1$ and any size function $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ with $f(k) = O(k \log^{1-\varepsilon} k)$, the integrality gap of the LP-relaxation (*LP-Metric*) is $\mathsf{IG}_f(k) = \Omega(\varepsilon \log \log k)$.*

We remark that our lower bound for the integrality gap of 0EwSN(f) does not imply a size lower bound for $O(\log \log k)$-quality flow sparsifiers. However, if the same lower bound can be proved for a slightly generalized version of 0EwSN(f), that was proposed in [1] and analyzed in [12], then it will imply an $\Omega(k \log^{1-\varepsilon} k)$ size lower bound for $O(\log \log k)$-quality flow sparsifiers. We provide a detailed discussion in Appendix A.

1.2 Technical Overview

We now discuss some high-level ideas in the proof of Theorem 1. Given any k, we will construct an unweighted graph G on n vertices (where $n \approx k \log k$) and k terminals, and show that any solution of the 0EwSN problem with size $O(k \log^{1-\varepsilon} k)$ has cost lower bounded by $\Omega(\log \log k)$ times the number of edges in G.

Our hard instance is a constant degree expander (with an arbitrary set of its k vertices as terminals). There are two main reasons to choose such a graph. First, in previous work [8] for proving the $\Omega(\sqrt{\log k})$ integrality gap lower bound for the 0-Extension problem, a graph called "expander with tails" was used. Though the tails in their construction appear useless for our purpose, as we allow a super-linear number of Steiner vertices which easily accomodate a single-edge tail for each terminal, the expander graph turns out to still be the core structure that is hard to compress. Second, it has been shown [3] that 0-Extension problem on minor-free graphs has integrality gap $O(1)$, so our hard example has to contain large cliques as minor, which makes expanders, a favorable choice. For technical reasons, we need some additional properties like Hamiltonicity and high girth. See Section 3.1 for more details.

Next we want to lower bound the cost of an 0EwSN solution of size $O(k \log^{1-\varepsilon} k)$. We first consider a special type of solutions where each cluster is mapped to a vertex on the graph. Recall that a solution consists of a partition $\mathcal{F}$ of $V(G)$ into clusters and a metric δ on clusters in $\mathcal{F}$. Specifically, in this special type of solutions, we require that each cluster $F \in \mathcal{F}$ corresponds to a distinct vertex v_F (called its *center*) in G, and for all pairs F, F' the metric $\delta(F, F')$ coincides with the shortest-path distance between $v_F, v_{F'}$ in G. We show that all special solutions have cost $\Omega(n \cdot \varepsilon \log \log k)$. Intuitively, as the graph is a constant degree expander, every center v_F is within distance $(\varepsilon/100) \cdot \log \log k$ to at most $(\log k)^{\varepsilon/10}$ other centers, but its cluster F contains $(\log k)^{\varepsilon}$ vertices on average and so it has $\Omega((\log k)^{\varepsilon})$ inter-cluster edges. As we measure the distance between clusters using their centers, only a small fraction of the inter-cluster edges will cost less than $(\varepsilon/100) \cdot \log \log k$, making the total new edge length $\Omega(n \cdot \varepsilon \log \log k)$ (as the number of inter-cluster edges in a balanced expander partition is $\Omega(n)$). Careful calculations are needed to turn these informal arguments into a rigorous proof. See Section 3.2 for more details.

Afterwards, we show that general solutions can actually be reduced to the special type of solutions considered in the first step, losing only an $O(1)$ factor in its cost. In fact, it has been recently shown [12] that, to analyze the cost of any 0EwSN instance, it suffices to consider 0EwSN solutions whose metric δ is embeddable into a geodesic structure of the terminal-induced shortest path distance metric called *tight span*. Our main contribution here, on a conceptual level, is showing that *for a graph with high girth, its tight span structure locally coincides with the graph structure itself*. In this sense, we can compare any solution to some special solution considered in Step 1. On a global level, in such a comparison it turns out that we will lose a factor which is approximately the *diameter-girth ratio*, which we can manage to get to $O(1)$ with an additional short-cycle-removing step in the construction of the expander. We believe this diameter-girth ratio quantifies how "local" a graph structure is, and should be of independent interest to other graph problems on shortest-path distances. To carry out the technical steps, we employ a notion called *continuiazation* of a graph recently studied in [11]. See Section 3.3 for more details.

2 Preliminaries

By default, all logarithms are to the base of 2.

Let $G = (V, E, \ell)$ be an edge-weighted graph, where each edge $e \in E$ has weight (or *length*) ℓ_e. For a vertex $v \in V$, we denote by $\deg_G(v)$ the degree of v in G. For each pair $S, T \subseteq V$ of disjoint subsets, we denote by $E_G(S, T)$ the set of edges in G with one endpoint in S the other endpoint in T. For a pair v, v' of vertices in G, we denote by $\mathsf{dist}_G(v, v')$ (or $\mathsf{dist}_\ell(v, v')$) the shortest-path distance between v and v' in G. We define the *diameter* of G as $\mathsf{diam}(G) = \max_{v,v' \in V} \{\mathsf{dist}_G(v, v')\}$, and we define the *girth* of G, denoted by $\mathsf{girth}(G)$, as the minimum weight of any cycle in G. We may omit the subscript G in the above notations when the graph is clear from the context.

Given a graph G, its *conductance* is defined as

$$\Phi(G) = \min_{S \subseteq V, S \neq \emptyset, S \neq V} \left\{ \frac{|E_G(S, V \setminus S)|}{\min\left\{\sum_{v \in S} \deg_G(v), \sum_{v \notin S} \deg_G(v)\right\}} \right\}.$$

We say that G is a *ϕ-expander* iff $\Phi(G) \geq \phi$.

3 Proof of Theorem 1

In this section we prove the main result Theorem 1, which shows that, when we only have $O(k \log^{1-\varepsilon} k)$ Steiner nodes, the best ratio we can get is $\Omega(\varepsilon \log \log k)$. We begin by describing the hard instance in Section 3.1, which is essentially a high-girth expander with a subset of vertices designated as terminals. Then in Section 3.2 we show that a special type of solutions may not have small cost. Finally, in Section 3.3 we generalize the arguments in Section 3.2 to analyze an arbitrary solution, completing the proof of Theorem 1. Some technical details in Section 3.3 are deferred to Section 3.4.

3.1 The Hard Instance

Let k be a sufficiently large integer. Let $n > k$ be an integer such that $k = \left\lceil \frac{n \log \log n}{\log n} \right\rceil$. Let V be a set of n vertices. Let Σ be the set of all permutations on V. For a permutation $\sigma \in \Sigma$, we define its corresponding edge set $E_\sigma = \{(v, \sigma(v)) \mid v \in V\}$.

We now define the hard instance (G, T, D). Graph G is constructed in two steps. In the first step, we construct an auxiliary graph G'. Its vertex set is V. Its edge set is obtained as follows. We sample three permutations $\sigma_1, \sigma_2, \sigma_3$ uniformly at random from Σ, and then let $E(G') = E_{\sigma_1} \cup E_{\sigma_2} \cup E_{\sigma_3}$. In the second step, we remove all short cycles in G' to obtain G. Specifically, we first compute a breath-first-search tree τ starting from an arbitrary vertex of G'. We then iteratively modify G' as follows. While G' contains a cycle C of length at most $(\log n)/100$, we find an edge of $C \setminus \tau$ (note that such an edge must exist, as τ is a tree), and remove it from G'. We continue until G' no longer contains cycles of length at most $(\log n)/100$. We denote by G the resulting graph. The terminal set T is an arbitrary subset of V with size k. For each edge $e \in E(G)$, its weight $c(e)$ is defined to be 1, and its *length* ℓ_e is also defined to be 1. The metric D on the set T of terminals is simply defined to be the shortest-path distance (in G) metric on T induced by edge length $\{\ell_e\}_{e\in E(G)}$.

We next show some basic properties of the graphs G' and G. We start with the following observations and propositions.

▶ **Observation 2.** *G' is a 6-regular graph, so $|E(G)| \le |E(G')| \le 3n$.*

▶ **Observation 3.** $\mathsf{girth}(G) \ge (\log n)/100$.

▶ **Proposition 4.** *The probability that $|E(G') \setminus E(G)| \ge n^{0.3}$ is at most $O(n^{-0.2})$.*

Proof. Let $v_1, \dots, v_L$ be a sequence of $L \le (\log n)/100$ distinct vertices of V. We now show that the probability that the cycle $(v_1, \dots, v_L, v_1)$ exists in $E(G')$ is at most $\big(6/(n-L)\big)^L$. Indeed, to realize the cycle edge (v_i, v_{i+1}), for some $\ell \in \{1,2,3\}$, $\sigma_\ell(v_i) = v_{i+1}$ or $\sigma_\ell(v_{i+1} = v_i$, where for convenience we say $v_{L+1} = v_1$. There are 6 possible events. In order to form the cycle, we need to form L edges, and each edge has 6 possible events, which means there are at most 6^L ways to form the cycle in total. Consider any possible way, we have $\ell_1 \dots, \ell_L \in \{1,2,3\}$ and $j_1, \dots, j_L \in \{0,1\}$ such that for any index $1 \le i \le L$, we have $\sigma_{\ell_i}(v_{i+j_i}) = v_{i+1-j_i}$. Let $\mathcal{E}_i$ denote this event, we have $\Pr[\mathcal{E}_i|\mathcal{E}_1, \dots, \mathcal{E}_{i-1}] \le 1/(n-i)$. Thus the probability that all events $\mathcal{E}_i$ happen is at most $1/(n-L)^L$. Applying union bound on all the ways to form the cycle, the probability that the cycle exists in $E(G')$ is at most $\big(6/(n-L)\big)^L$.

Therefore, the expected number of cycles in G' with length at most $(\log n)/100$ is at most

$$\sum_{3\le L\le(\log n)/100} \frac{n(n-1)\cdots(n-L+1)}{2L\cdot\left(\frac{n-L}{6}\right)^L} \le \sum_{3\le L\le(\log n)/100} \frac{6^L}{2L}\cdot\left(1+\frac{L}{n-L}\right)^L \le \sum_{3\le L\le(\log n)/100} 6^L \le n^{0.1}.$$

Therefore, from Markov Bound, with probability $n^{-0.2}$, the number of cycles in G' with length at most $(\log n)/100$ is at most $n^{0.3}$. Note that we delete at most one edge per each cycle, so $|E(G') \setminus E(G)|$ is less than the number of cycles in G' with length at most $(\log n)/100$, the proposition follows. ◀

We use the following previous results on the conductance and the Hamiltonicity of G'.

▶ **Lemma 5** ([30]). *With probability $1 - o(1)$, $\Phi(G') = \Omega(1)$.*

▶ **Corollary 6.** *With probability $1 - o(1)$, the diameter of G is at most $O(\log n)$.*

Proof. From the construction of G, G contains a BFS tree of G', so the diameter of G is at most twice the diameter of G'. Therefore, it suffices to show that, if $\Phi(G') \ge \Omega(1)$, then the diameter of graph G' is at most $O(\log n)$, which we do next.

Let v be an arbitrary vertex of G'. For each integer t, we define the set $B_t = \{v' \mid \mathsf{dist}(v, v') \le t\}$, and $\alpha_t = \sum_{v':\mathsf{dist}_{G'}(v,v')\le t} \deg(v')$, namely the sum of degrees of all vertices in B_t.

Denote $t^* = \max\{t \mid \alpha_t \leq |E(G')|\}$. Note that, for each $1 \leq t \leq t^*$, as $\Phi(G') \geq \Omega(1)$, $|E(B_t, V \setminus B_t)| \geq \Omega(\alpha_t)$. Therefore,

$$\alpha_{t+1} \geq \alpha_t + \sum_{v' \in B_{t+1} \setminus B_t} \deg(v) \geq \alpha_t + |E(B_t, V \setminus B_t)| \geq \alpha_t \cdot (1 + \Omega(1)).$$

It follows that $t^* \leq O(\log n)$. Therefore, for any pair $v, v' \in V$, the set of vertices that are at distance at most $t^* + 1$ from v must intersect the set of vertices that are at distance at most $t^* + 1$ from v', as otherwise the sum of degrees in all vertices in these two sets is greater than $2|E(G')|$, a contradiction. Consequently, the diameter of G' is at most $2t^* + 2 \leq O(\log n)$. ◀

▶ **Lemma 7** ([18]). *With probability $1 - o(1)$, the subgraph of G' induced by edges of $E_{\sigma_1} \cup E_{\sigma_2}$ is Hamiltonian.*

Now if we consider the semi-metric LP relaxation (LP-Metric) of this instance (G, T, D), then clearly the graph itself gives a solution δ to (LP-Metric). Specifically, $\delta(u, u') = \mathsf{dist}_\ell(u, u')$, where $\mathsf{dist}_\ell(\cdot, \cdot)$ the shortest-path (in G) distance metric on V induced by the lengths $\{\ell_e\}_{e \in E(G)}$. Such a solution has cost $|E(G)| = O(n)$ (as all edges have weight $c(e) = 1$). Therefore, in order to prove Theorem 1, it suffices to show that any solution $(\mathcal{F}, \delta)$ with size $O(k \log^{1-\varepsilon} k)$ has cost at least $\Omega(\varepsilon n \log\log n) = \Omega(\varepsilon n \log\log k)$.

Observe that, the graph G constructed above is essentially a bounded-degree high-girth expander, which is similar to the hard instance used in [26] for proving the $\Omega(\log\log k)$ quality lower bound for flow vertex sparsifier (without Steiner nodes). However, our proof in the following subsections takes a completely different approach from the approach in [26].

3.2 Proof of Theorem 1 for Canonical Solutions

In this subsection, we prove the cost lower bound for a special type of solutions to the 0EwSN(f) instance (G, T, D) which we call canonical. Specifically, a solution is canonical $(\mathcal{F}, \delta)$ if

- each cluster $F \in \mathcal{F}$ corresponds to a distinct vertex of V, we call this vertex the *center* of F, denote as $v(F)$ (note however that $v(F)$ does not necessarily lie in F). For each terminal $t \in T$, the unique cluster $F \in \mathcal{F}$ that contains t, $v(F) = t$; and
- for each pair F, F' of clusters in $\mathcal{F}$, $\delta(F, F') = \mathsf{dist}_G(v(F), v(F'))$.

In this subsection, we show that, with high probability, any canonical solution of size $o(n/\log^\varepsilon n)$ has cost $\Omega(\varepsilon n \log\log n)$.

Consider now a canonical solution $(\mathcal{F}, \delta)$ to the instance. We say that $F \in \mathcal{F}$ is *large* iff $|F| \geq n^{0.1}$, otherwise we say it is *small*. We distinguish between the following cases, depending on the total size of large clusters.

Recall that $\mathsf{cost}(\mathcal{F}, \delta) = \sum_{(u,u') \in E(G)} \delta(F(u), F(u'))$, where $F(u)$ ($F(u')$, resp.) is the unique cluster in $\mathcal{F}$ that contains u (u', resp.). We call $\delta(F(u), F(u'))$ the *contribution* of edge (u, u') to the cost $\mathsf{cost}(\mathcal{F}, \delta)$.

Case 1: The total size of large clusters is at most $0.1n$

As the solution $(\mathcal{F}, \delta)$ is canonical,

$$\delta(F(u), F(u')) = \mathsf{dist}_G(v(F(u)), v(F(u')) \geq \mathsf{dist}_{G'}(v(F(u)), v(F(u')),$$

as G is obtained from G' by only deleting edges. We say that a pair F, F' of clusters are *friends* (denoted as $F \sim F'$), iff $\mathsf{dist}_{G'}(v(F), v(F')) \leq \varepsilon \log\log n/30$. We say that an edge

(u, u') is *unfriendly*, iff the pair of clusters that contain u and u' are not friends. Therefore, in order to show $\mathsf{cost}(\mathcal{F}, \delta) = \Omega(\varepsilon n \log\log n)$, it suffices to show that there are $\Omega(n)$ unfriendly edges in G'. In particular, since graph G is obtained from G' by deleting at most $n^{0.3}$ edges, there are $\Omega(n)$ edges contributing at least $\varepsilon \log\log n$ each to $\mathsf{cost}(\mathcal{F}, \delta)$. Note that, as G is a 6-regular graph, each cluster is a friend to at most $6^{\varepsilon \log\log n/30} < \log^{\varepsilon/10} n$ clusters in $\mathcal{F}$.

The following lemma shows that there with high probability, const fraction of the edges in G' are unfriendly edges whose lengths are $\Omega(\varepsilon \log\log n)$. This lemma completes the proof in this case.

▶ **Lemma 8.** *With probability $1-o(1)$, the random graph G' satisfies that, for any partition $\mathcal{F}$ of V into $|\mathcal{F}| \le O(n/\log^{\varepsilon} n)$ clusters such that $\sum_{|F|\ge n^{0.1}} |F| \le 0.1n$, and for any friendship relation on $\mathcal{F}$ in which each cluster F is a friend to at most $\log^{0.1\varepsilon} n$ other clusters, G' contains at least $n/10$ unfriendly edges, i.e., $\sum_{F\not\sim F'} |E_{G'}(F, F')| \ge n/10$.*

Proof. Recall that G' is obtained by sampling three random permutations $\sigma_1, \sigma_2, \sigma_3$ from Σ and taking the union of their corresponding edge sets $E_{\sigma_1}, E_{\sigma_2}, E_{\sigma_3}$. We alternatively view G' as constructed in two steps. In the first step, we obtain a graph $\hat{G}$ by sampling two random permutations σ_1, σ_2 from Σ and letting $\hat{G} = (V, E_{\sigma_1} \cup E_{\sigma_2})$. In the second step, we sample a third permutation σ_3 from Σ and let $G' = (V, E(\hat{G}) \cup E_{\sigma_3})$. From Lemma 7, with high probability, $\hat{G}$ contains a Hamiltonian cycle on V.

For convenience, we denote by $(\mathcal{F}, \sim)$ a pair of clustering $\mathcal{F}$ and the friendship relation on clusters of $\mathcal{F}$. We say that the pair $(\mathcal{F}, \sim)$ is *valid*, iff $|\mathcal{F}| \le O(n/\log^{\varepsilon} n)$, $\sum_{|F|\ge n^{0.1}} |F| \le 0.1n$, and each cluster F is a friend to at most $\log^{\varepsilon/10} n$ other clusters.

▷ Claim 9. For any Hamiltonian cycle C on V, there are at most $n^{n/4}$ valid pairs $(\mathcal{F}, \sim)$ satisfying that $\sum_{F\not\sim F'} |E_C(F, F')| < n/10$.

Proof. Denote by $L = c^* n/\log^{\varepsilon} n$ the number of clusters of $\mathcal{F}$, and let $\mathcal{F} = \{F_1, \ldots, F_L\}$.

First, the number of possible friendship relations on $\mathcal{F}$ such that each cluster of $\mathcal{F}$ is a friend to at most $\log^{0.1\varepsilon} n$ other clusters is at most

$$\binom{L}{\log^{0.1\varepsilon} n}^{L} \le \binom{\frac{c^* n}{\log^{\varepsilon} n}}{\log^{0.1\varepsilon} n}^{\frac{c^* n}{\log^{\varepsilon} n}} \le \left(\frac{c^* n}{\log^{\varepsilon} n}\right)^{\frac{c^* n}{\log^{\varepsilon} n}\cdot \log^{0.1\varepsilon} n} < n^{c^* n \log^{-0.05\varepsilon} n}.$$

Assume now that we have a fixed friendship relation $\sim$ on the clusters in $\mathcal{F}$. We now count the number of clusterings $\mathcal{F}$ with $\sum_{F\not\sim F'} |E_C(F, F')| < n/10$. Denote $C = (v_1, v_2, \ldots, v_n, v_1)$. First, the number of possible unfriendly edge set (which is a subset of $E(C)$ of size at most $0.1n$) is at most

$$\sum_{i=0}^{n/10} \binom{n}{i} \le n \cdot \binom{n}{n/10} < n \cdot \left(\frac{en}{n/10}\right)^{n/10} < n^{n \log^{-0.5} n}.$$

We now count the number of clusterings $\mathcal{F}$ that, together with the fixed friendship relation $\sim$, realizes a specific unfriendly edge set. We will sequentially pick, for each i from 1 to n, a set among $\{F_1, \ldots, F_L\}$ to add the vertex v_i to. The first vertex v_1 has L choices. Consider now some index $1 \le i \le n-1$ and assume that we have picked sets for $v_1, \ldots, v_i$. If (v_i, v_{i+1}) is an unfriendly edge, then vertex v_{i+1} has L choices; if (v_i, v_{i+1}) is not an unfriendly edge, this means that v_{i+1} must go to some cluster that is a friend of the cluster we have picked for v_i (or v_{i+1} can go to the same cluster as v_i), so v_{i+1} has at most $\log^{0.1\varepsilon} n + 1$ choices. As there are no more than $0.1n$ unfriendly edges, the number of possible clusterings $\mathcal{F}$ is at most

$$n \cdot (\log^{0.1\varepsilon} n)^n \cdot \left(\frac{n}{\log^{\varepsilon} n}\right)^{0.1n} < n^{n/5}.$$

Altogether, the number of valid pairs $(\mathcal{F}, \sim)$ satisfying that $\sum_{F \not\sim F'} |E_C(F, F')| \geq n/10$ is at most

$$n^{c^* n \log^{-0.05\varepsilon} n} \cdot n^{n \log^{-0.5} n} \cdot n^{n/5} < n^{n/4}. \quad \triangleleft$$

▷ Claim 10. For every valid pair $(\mathcal{F}, \sim)$, the probability that the edge set E_{σ_3} of a random permutation σ_3 contains at most $n/10$ unfriendly edges is at most $n^{-n/3}$.

Proof. We say that a cluster $F \in \mathcal{F}$ is *bad* if it does not have a friend cluster of size at least $n^{0.4}$, otherwise we say it is *good*. We first prove the following observation that most vertices lie in a bad cluster.

▶ **Observation 11.** $\sum_{F:bad} |F| \geq 0.8n$.

Proof. As the pair $(\mathcal{F}, \sim)$ is valid, $\sum_{|F| \geq n^{0.1}} |F| \leq 0.1n$, so $\mathcal{F}$ contains at most $0.1 \cdot n^{0.6}$ clusters with size at least $n^{0.4}$. As each cluster is a friend to at most $\log^{0.1\varepsilon} n$ other clusters, $\mathcal{F}$ contains at most $\left(0.1 \cdot n^{0.6} \cdot \log^{0.1\varepsilon} n\right)$ good sets. Therefore, the total size of all good clusters is at most $0.1n + n^{0.1} \cdot 0.1 \cdot n^{0.6} \cdot \log^{0.1\varepsilon} n < 0.2n$. The observation now follows. ◀

We alternatively construct the random permutation σ_3 as follows. We arrange the vertices in V into a sequence $(v_1, \ldots, v_n)$, such that each of the first half $v_1, \ldots, v_{n/2}$ lies in some bad set. Now sequentially for each $1, 2, \ldots, n$, we sample a vertex u_i (without replacement) from V and designate it as $\sigma_3(v_i)$. It is easy to observe that the permutation σ_3 constructed in this way is a random permutation from Σ.

The following observation completes the proof of Claim 10.

▶ **Observation 12.** *The probability that the number of unfriendly edges in* $\{(v_i, \sigma_3(v_i)) \mid 1 \leq i \leq 9n/10\}$ *is less than* $0.1n$ *is at most* $n^{-n/3}$.

Proof. For any v in a bad cluster, the number of vertices in its friend clusters is at most $n^{0.4} \log^{0.1} n$. For each $1 \leq i \leq 9n/10$, when we pick $\sigma_3(v_i)$, we have at least $n/10$ choices from the remaining element in V, and as v_i is in a bad set, at most $n^{0.4} \log^{0.1} n$ of them will not create an unfriendly edge. Therefore, the probability that the edge we sample is not a bad edge is at most $\frac{1}{\sqrt{n}}$. Let X_i be the indicator random variable such that $X_i = 1$ if $(v_i, \sigma_3(v_i))$ is not a bad edge. By Azuma's Inequality (Chernoff Bounds on martingales, see e.g., [25]),

$$\Pr\left[\sum_{i=1}^{2n/3} X_i > 4n/5\right] < \left(\frac{5}{4\sqrt{n}}\right)^{4n/5} < n^{-n/3}.$$

Thus, with probability at least $1 - n^{-n/3}$, the set $\{(v_i, \sigma_3(v_i)) | 1 \leq i \leq 9n/10\}$ contains at least $9n/10 - 4n/5 = n/10$ bad edges. ◀

◁

Combining Claim 9 and Claim 10, we get that, over the randomness in the construction of G', the probability that there exists a pair $(\mathcal{F}, \sim)$ in which each cluster F is a friend to at most $\log^{0.1\varepsilon} n$ other clusters, such that G' contains less than $n/10$ unfriendly edges, is at most $n^{-n/3} \cdot n^{n/4} = n^{-n/12}$. This completes the proof of Lemma 8. ◀

Case 2: The total size of large clusters is greater than $0.1n$

We denote by V' the union of all large clusters in $\mathcal{F}$. We start by proving the following claim.

▷ Claim 13. There exists a collection of $k/4$ edge-disjoint paths in G, such that each path connects a distinct terminal to a distinct vertex of V'.

Proof. We construct a graph $\hat{G}$ as follows. We start from graph G', and add two vertices s,t to it. We then connect s to each terminal in T by an edge, and connect each vertex in V' to t by an edge. All edges in $\hat{G}$ has unit capacity. We claim that there exists a collection $\mathcal{P}$ of $k/3$ edge-disjoint paths in $\hat{G}$, such that each path connects a distinct terminal to a distinct vertex of V'. Note that this implies Claim 13. This is because the number of edges in $E(G')\setminus E(G)$ is at most $n^{0.3} < k/12$, and each such edge is contained in at most one path in $\mathcal{P}$ (since the paths in $\mathcal{P}$ are edge-disjoint), so at least $k/3 - k/12 \geq k/4$ paths in $\mathcal{P}$ are entirely contained in G. We now prove the claim. From the max-flow min-cut theorem, it suffices to show that the minimum s-t cut in $\hat{G}$ contains at least $k/3$ edges.

Consider any s-t cut $(S\cup\{s\}, (V\setminus S)\cup\{t\})$ in $\hat{G}$ and denote by E' the set of edges in this cut. We distinguish between the following cases.

Case 1: $|S| \leq |V|/2$. Recall that G is a 6-regular graph, so $\sum_{v\in S}\deg(v) \leq \sum_{v\notin S}\deg(v)$. Then from Lemma 5, $|E'| \geq \sum_{v\in S}\deg(v)/2 \geq |S|/2$. If $|S| \geq 2k/3$, then $|E'| \geq k/3$. If $|S| < 2k/3$, then at least $k/3$ terminals lie in $V\setminus S$. As there is an edge connecting s to each terminal, $|E'| \geq k/3$.

Case 2: $|S| > |V|/2$. Via similar arguments, we can show that $|E'| \geq |V'|/3 \geq 0.1n/3 > k/3$. ◁

We denote by $\mathcal{P}$ the collection of paths given by Claim 13. We now use these paths to complete the proof. Consider such a path $P = (u_1,\dots,u_r)$. Denote by F_i the cluster that contains u_i, then the contribution of P to the cost $\mathsf{cost}(\mathcal{F},\delta)$ is

$$\sum_{(u_i,u_{i+1})\in E(P)} \delta(F_i,F_{i+1}) = \sum_{1\leq i\leq r-1} \mathsf{dist}_G(v(F_i),v(F_{i+1})) \geq \mathsf{dist}_G(v(F_1),v(F_r)) \geq \mathsf{dist}_{G'}(v(F_1),v(F_r)).$$

(We have used the property that for every pair $v,v'\in V$, $\mathsf{dist}_G(v,v') \geq \mathsf{dist}_{G'}(v,v')$, as G is obtained from G' by only deleting edges.)

Recall P connects a terminal to a vertex in V'. Recall that each large cluster has size at least $n^{0.1}$, so there are at most $n^{0.9}$ of them. Therefore, if we denote by V'' the subset of vertices that large clusters corresponds to, then $|V''| \leq n^{0.9}$. For each path $P\in\mathcal{P}$, we denote by t_P the terminal endpoint of P (that is, $u_1 = v(F_1) = t_P$), and by v''_P the vertex that the cluster containing u_r corresponds to (that is, $v''_P = v(F_r)$), then $\sum_{(u_i,u_{i+1})\in E(P)}\delta(F_i,F_{i+1}) \geq \mathsf{dist}_\ell(t_P,v''_P)$. As the paths in $\mathcal{P}$ are edge-disjoint, their contribution to $\mathsf{cost}(\mathcal{F},\delta)$ can be added up, i.e.,

$$\mathsf{cost}(\mathcal{F},\delta) \geq \sum_{P\in\mathcal{P}} \mathsf{dist}_\ell(t_P,v''_P). \tag{1}$$

On the one hand, as graph G' is 6-regular, for each $v''\in V''$, the number of vertices at distance at most $\log n/100$ to v'' is at most $6^{\log n/100} \leq n^{1/30}$. Therefore, there are at most $n^{1/30}\cdot n^{0.9} = n^{14/15}$ terms on the RHS of Equation (1) that at most $\leq \log n/100$. On the other hand, there are at least $k/4 = \Omega(\frac{n\log\log n}{\log n})$ terms on the RHS of Equation (1), so at least $k/4 - n^{14/15} \geq k/5$ terms has value at least $\log n/100$. Consequently, $\mathsf{cost}(\mathcal{F},\delta) \geq (k/5)\cdot(\log n/100) = \Omega(k\log n) = \Omega(n\log\log n)$.

3.3 Completing the Proof of Theorem 1

We have shown in Section 3.2 all canonical solutions with size $O(k\log^{1-\varepsilon} n)$ have cost $\Omega(\varepsilon n \log\log n)$. In this subsection, we complete the proof of Theorem 1 by showing that, intuitively, an arbitrary solution $(\mathcal{F},\delta)$ to the instance (G,T,D) can be "embedded" into a canonical solution, without increasing its cost by too much.

We start by introducing the notion of *continuization*.

Continuization of a graph

Let $G=(V,E,\ell)$ be an edge-weighted graph. Its *continuization* is a metric space $(V^{\mathsf{con}},\ell^{\mathsf{con}})$, that is defined as follows. Each edge $(u,v)\in E$ is viewed as a continuous line segment $\mathsf{con}(u,v)$ of length $\ell_{(u,v)}$ connecting u,v, and the point set V^{con} is the union of the points on all lines $\{\mathsf{con}(u,v)\}_{(u,v)\in E}$. Specifically, for each edge $(u,v)\in E$, the line $\mathsf{con}(u,v)$ is defined as

$$\mathsf{con}(u,v)=\left\{(u,\alpha)\mid 0\le\alpha\le\ell_{(u,v)}\right\}=\left\{(v,\beta)\mid 0\le\beta\le\ell_{(u,v)}\right\},$$

where (u,α) refers to the unique point on the line that is at distance α from u, and (v,β) refers to the unique point on the line that is at distance β from v, so $(u,\alpha)=(v,\ell_{(u,v)}-\alpha)$.

The metric ℓ^{con} on V^{con} is naturally induced by the shortest-path distance metric $\mathsf{dist}_\ell(\cdot,\cdot)$ on V as follows. For a pair p,p' of points in V^{con},

- if p,p' lie on the same line (u,v), say $p=(u,\alpha)$ and $p'=(u,\alpha')$, then $\ell^{\mathsf{con}}(p,p')=|\alpha-\alpha'|$;
- if p lies on the line (u,v) with $p=(u,\alpha)$ and p' lies on the line (u',v') with $p'=(u',\alpha')$, then

$$\begin{aligned}\ell^{\mathsf{con}}(p,p')=\min\{&\mathsf{dist}_\ell(u,u')+\alpha+\alpha',\quad \mathsf{dist}_\ell(u,v')+\alpha+(\ell_{(u',v')}-\alpha'),\\ &\mathsf{dist}_\ell(v,u')+(\ell_{(u,v)}-\alpha)+\alpha',\quad \mathsf{dist}_\ell(v,v')+(\ell_{(u,v)}-\alpha)+(\ell_{(u',v')}-\alpha')\}.\end{aligned}$$

Clearly, every vertex $u\in V$ also belongs to V^{con}, and for every pair $u,u'\in V$, $\mathsf{dist}_\ell(u,u')=\ell^{\mathsf{con}}(u,u')$. For a path P in G connecting u to u', it naturally corresponds to a set P^{con} of points in V^{con}, which is the union of all lines corresponding to edges in $E(P)$. The set P^{con} naturally inherits the metric ℓ^{con} restricted on P^{con}. We will also call P^{con} a path in the continuization $(V^{\mathsf{con}},\ell^{\mathsf{con}})$.

We show that, for each graph G with a set T of terminals, then any other metric w on a set of points containing T, such that w restricted on T is identical to dist_G restricted on T, can be "embedded" into the continuation of G, with expected stretch bounded by some structural measure that only depends on G. Specifically, we prove the following main technical lemma.

▶ **Lemma 14.** *Let (G,T,ℓ) be any instance of* 0EwSN *such that G is not a tree (so* $\mathsf{girth}(G)<+\infty$*), and let $(\mathcal{F},\delta)$ be any solution to it. Let $(V^{\mathsf{con}},\ell^{\mathsf{con}})$ be the continuization of graph G. Then there exists a random mapping $\phi:\mathcal{F}\to V^{\mathsf{con}}$, such that*

- *for each terminal $t\in T$, if F is the (unique) cluster in $\mathcal{F}$ that contains t, then $\phi(F)=t$; and*
- *for every pair $F,F'\in\mathcal{F}$,*

$$\mathbb{E}\left[\ell^{\mathsf{con}}(\phi(F),\phi(F'))\right]\le O\left(\frac{\mathsf{diam}(G)}{\mathsf{girth}(G)}\right)\cdot\delta(F,F').$$

Before we prove Lemma 14 in Section 3.4, we provide the proof of Theorem 1 using it.

Proof of Theorem 1. Consider any solution $(\mathcal{F},\delta)$ to the instance (G,T,ℓ) constructed in Section 3.1 with size $|\mathcal{F}|\le o(k\cdot\log^{1-\varepsilon}k)$. From Observation 3 and Corollary 6, $\frac{\mathsf{diam}(G)}{\mathsf{girth}(G)}\le\frac{O(\log n)}{(\log n)/100}=O(1)$. From Lemma 14, there exists a random mapping $\phi:\mathcal{F}\to V^{\mathsf{con}}$, such that for every pair $F,F'\in\mathcal{F}$, $\mathbb{E}\left[\ell^{\mathsf{con}}(\phi(F),\phi(F'))\right]\le O(1)\cdot\delta(F,F')$.

Fix such a mapping ϕ, we define a canonical solution $(\mathcal{F},\hat{\delta})$ based on $(\mathcal{F},\delta)$ as follows. The collection of clusters is identical to the collection $\mathcal{F}$. For each $F\in\mathcal{F}$, recall that $\phi(F)$ is a point in V^{con}. Assume the point $\phi(F)$ lies on the line (u,v) and is closer to u than to v (i.e., $\ell^{\mathsf{con}}(\phi(F),u)\le\ell^{\mathsf{con}}(\phi(F),v)$), then we let u be the vertex in V that it corresponds to. For each pair $F,F'\in\mathcal{F}$, with F corresponding to u_F and and F' corresponding to $u_{F'}$, we define $\hat{\delta}(F,F')=\mathsf{dist}_\ell(u_F,u_{F'})$. As graph G in the instance (G,T,δ) constructed in Section 3.1 is an unweighted graph, it is easy to see that

$$\hat{\delta}(F,F')=\mathsf{dist}_\ell(u_F,u_{F'})\le\ell^{\mathsf{con}}(\phi(F),\phi(F'))+2.$$

As the mapping ϕ is random, $\hat{\delta}$ is also random, and so $\mathbb{E}\left[\hat{\delta}(F,F')\right]\le O(1)\cdot\delta(F,F')+2$.

From the properties of mapping ϕ in Lemma 14, we are guaranteed that such a solution $(\mathcal{F},\hat{\delta})$ is a canonical solution. Moreover, from linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[\mathsf{cost}(\mathcal{F},\hat{\delta})\right]=\mathbb{E}\left[\sum_{(u,v)\in E}\hat{\delta}(F(u),F(v))\right]&=\sum_{(u,v)\in E}O\Big(\delta(F(u),F(v))\Big)+2\\&=O\Big(\mathsf{cost}(\mathcal{F},\delta)\Big)+O(n).\end{aligned}$$

Therefore, it follows that there exists a canonical solution $(\mathcal{F},\hat{\delta})$, such that $\mathsf{cost}(\mathcal{F},\hat{\delta})\le O(\mathsf{cost}(\mathcal{F},\delta)+n)$. As we have shown in Section 3.2 that any canonical solution $(\mathcal{F},\hat{\delta})$ with $|\mathcal{F}|\le o(k\log^{1-\varepsilon}k)$ satisfies that $\mathsf{cost}(\mathcal{F},\hat{\delta})=\Omega(\varepsilon n\log\log n)$, it follows that $\mathsf{cost}(\mathcal{F},\delta)=\Omega(\varepsilon n\log\log n)$. This implies that the integrality gap of (LP-Metric) is at least $\Omega(\varepsilon\log\log n)$. ◀

3.4 Proof of Lemma 14

In this subsection, we provide the proof of Lemma 14. We first consider the special case where G is a tree, and then prove Lemma 14 for the general case.

▶ **Lemma 15.** *Let (G,T,ℓ) be an instance of* 0EwSN *where G is a tree. Let $(\mathcal{F},\delta)$ be an solution to it. Let $(V^{\mathsf{con}},\ell^{\mathsf{con}})$ be the continuization of G. Then there exists a mapping $\phi:\mathcal{F}\to V^{\mathsf{con}}$, such that*

- *for each terminal $t\in T$, if F is the (unique) cluster in $\mathcal{F}$ that contains t, then $\phi(F)=t$; and*
- *for every pair $F,F'\in\mathcal{F}$, $\ell^{\mathsf{con}}(\phi(F),\phi(F'))\le\delta(F,F')$.*

Proof. For each terminal $t\in T$, we denote by F_t the cluster in $\mathcal{F}$ that contains it, and set $\phi(F_t)=t$. For each cluster $F\in\mathcal{F}$ that does not contain any terminals, we define

$$\nu(F)=\min\left\{\frac{1}{2}\cdot\big(\delta(F,F_t)+\delta(F,F_{t'})-\delta(F_t,F_{t'})\big)\mid t,t'\in T\right\}.$$

Denote by t_1,t_2 the pair of terminals (t,t') that minimizes $(\delta(F,F_t)+\delta(F,F_{t'})-\delta(F_t,F_{t'}))/2$. As G is a tree, there is a unique shortest path connecting t_1 to t_2 in G, and therefore there exists a unique point p in V^{con} (that lies on the t_1-t_2 shortest path in V^{con}) with $\ell^{\mathsf{con}}(p,t_1)=\delta(F,F_{t_1})-\nu(F)$ and $\ell^{\mathsf{con}}(p,t_2)=\delta(F,F_{t_2})-\nu(F)$. We set $\phi(F)=p$.

▷ Claim 16. For every cluster $F \in \mathcal{F}$ and every terminal $t \in T$, $\ell^{\mathsf{con}}(\phi(F), t) \le \delta(F, F_t) - \nu(F)$.

Proof. Note that the point $\phi(F)$ lies on the (unique) shortest path between a pair t_1, t_2 of terminals. For $t \in \{t_1, t_2\}$, clearly the claim holds. Consider any other terminal t. Clearly $\phi(F)$ lies on either the path connecting t to t_1 or the path connecting t to t_2. Assume without lose of generality that $\phi(F)$ is on the path connecting t_1 and t. Since G is a tree, $\ell^{\mathsf{con}}(\phi(F), t_1) + \ell^{\mathsf{con}}(\phi(F), t) = \ell^{\mathsf{con}}(t_1, t) = \mathsf{dist}_\ell(t_1, t)$. On the other hand, by definition of $\nu(F)$ and $\phi(F)$, $\delta(F, F_{t_1}) + \delta(F, F_t) \ge \delta(F_{t_1}, F_t) + 2 \cdot \nu(F)$ holds, and $\ell^{\mathsf{con}}(\phi(F), t_1) = \delta(F, F_{t_1}) - \nu(F)$. Therefore, $\delta(F, F_t) \ge \ell^{\mathsf{con}}(\phi(F), t) + \nu(F)$. ◁

We now show that, for every pair $F, F' \in \mathcal{F}$, $\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le \delta(F, F')$. Denote by t_1, t_2 the pair of terminals whose shortest path contains $\phi(F)$, and by t'_1, t'_2 the pair of terminals whose shortest path contains $\phi(F')$. Assume without lose of generality that $\phi(F)$ is on the tree path between $\phi(F')$ and t_1, so $\ell^{\mathsf{con}}(\phi(F'), \phi(F)) + \ell^{\mathsf{con}}(\phi(F), t_1) = \ell^{\mathsf{con}}(\phi(F'), t_1)$. On the other hand, from the definition of $\phi(F)$ and Claim 16, $\ell^{\mathsf{con}}(\phi(F), t_1) = \delta(F, F_{t_1}) - \nu(F)$ and $\ell^{\mathsf{con}}(\phi(F'), t_1) \le \delta(F', F_{t_1}) - \nu(F')$. Therefore,

$$\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le \delta(F', F_{t_1}) - \delta(F, F_{t_1}) - \nu(F') + \nu(F) \le \delta(F, F') + \nu(F) - \nu(F').$$

Similarly, $\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le \delta(F, F') + \nu(F') - \nu(F)$. Altogether, $\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le \delta(F, F')$. ◀

In the remainder of this subsection, we complete the proof of Lemma 14. Denote $g = \mathsf{girth}(G)$. Let r be a real number chosen uniformly at random from the interval $[g/60, g/30]$, so $r \le g/30$ always holds. For each terminal $t \in T$, we denote by F_t the cluster in $\mathcal{F}$ that contains it, and set $\phi(F_t) = t$, so the first condition in Lemma 14 is satisfied.

For each cluster $F \in \mathcal{F}$, we define $A_F = \min\{\delta(F, F_t) \mid t \in T\}$. We first determine the image $\phi(F)$ for all clusters F with $A_F \le r$, in a similar way as Lemma 15 as follows.

Define

$$\nu(F) = \min\left\{\frac{1}{2} \cdot \left(\delta(F, F_t) + \delta(F, F_{t'}) - \frac{\delta(F_t, F_{t'})}{2}\right) \mid t, t' \in T\right\}.$$

Denote by t_1, t_2 the pair (t, t') that minimizes the above formula. We prove the following claim.

▷ Claim 17. $\delta(F, F_{t_1}) + \delta(F, F_{t_2}) \le 4 \cdot A_F \le 4r$.

Proof. Let t be the terminal such that $\delta(F, F_t) = A_F$. By definition of $\nu(F)$,

$$\nu(F) \le \frac{\delta(F, F_t) + \delta(F, F_t) - \frac{1}{2} \cdot \delta(F_t, F_t)}{2} = A_F.$$

On the other hand, from triangle inequality,

$$\nu(F) = \frac{\delta(F, F_{t_1}) + \delta(F, F_{t_2}) - \frac{1}{2} \cdot \delta(F_{t_1}, F_{t_2})}{2} \ge \frac{\delta(F, F_{t_1}) + \delta(F, F_{t_2})}{4}.$$

Altogether, $\delta(F, F_{t_1}) + \delta(F, F_{t_2}) \le 4 \cdot A_F$. ◁

From Claim 17, $\delta(F_{t_1}, F_{t_2}) \le \delta(F, F_{t_1}) + \delta(F, F_{t_2}) \le 4r < g/3$. Therefore, there is a unique shortest path connecting t_1 to t_2 in G, as otherwise G must contain a cycle of length at most $2g/3$, contradicting the fact that $\mathsf{girth}(G) = g$.

We then set $\phi(F)$ to be the point in V^{con} that lies in the t_1-t_2 shortest path, such that $\ell^{\mathsf{con}}(\phi(F),t_1) = 2\cdot(\delta(F,F_{t_1})-\nu(F))$ and $\ell^{\mathsf{con}}(\phi(F),t_2) = 2\cdot(\delta(F,F_{t_2})-\nu(F))$. Note that, by definition of t_1,t_2,

$$\ell^{\mathsf{con}}(\phi(F),t_1)+\ell^{\mathsf{con}}(\phi(F),t_2) = 2\cdot\delta(F,F_{t_1})+2\cdot\delta(F,F_{t_2})-4\cdot\nu(F) = \delta(F_{t_1},F_{t_2}).$$

We prove the following claim, that is similar to Claim 16.

▷ Claim 18. For every terminal t with $\delta(F,F_t)\le 6r$, $\ell^{\mathsf{con}}(\phi(F),t)\le 2\cdot(\delta(F,F_t)-\nu(F))$.

Proof. From Claim 17,

$$\begin{aligned}\delta(F_t,F_{t_1})+\delta(F_t,F_{t_2})+\delta(F_{t_1},F_{t_2}) &\le \Big(2\cdot\delta(F,F_t)+\delta(F,F_{t_1})+\delta(F,F_{t_2})\Big)+\delta(F_{t_1},F_{t_2})\\ &\le 12r+4r+4r+4r<g.\end{aligned}$$

Therefore, the point $\phi(F)$ must lie on either the t-t_1 shortest path or the t-t_2 shortest path in V^{con}, as otherwise the union of t-t_1 shortest path, t-t_2 shortest path, and t_1-t_2 shortest path contains a cycle of length less than g, a contradiction.

Assume without loss of generality that $\phi(F)$ lies on the t-t_1 shortest path, so

$$\ell^{\mathsf{con}}(\phi(F),t) = \delta(F_t,F_{t_1})-\ell^{\mathsf{con}}(\phi(F),t_1) = \delta(F_t,F_{t_1})-2(\delta(F,F_{t_1})-\nu(F)).$$

By definition of $\nu(F)$, $\delta(F_t,F_{t_1})\le 2\cdot(\delta(F,F_t)+\delta(F,F_{t_1})-2\cdot\nu(F))$. Therefore,

$$\ell^{\mathsf{con}}(\phi(F),t)\le 2\bigg(\delta(F,t)+\delta(F,F_{t_1})-2\nu(F)-\big(\nu(F,F_{t_1})-\nu(F)\big)\bigg) = 2(\delta(F,F_t)-\nu(F)).$$

◁

We now show in the next claim that the second condition in Lemma 14 holds for pairs of clusters that are close in δ.

▷ Claim 19. For every pair $F,F'\in\mathcal{F}$ with $A_F,A_{F'},\delta(F,F')\le r$, $\ell^{\mathsf{con}}(\phi(F),\phi(F'))\le 2\cdot\delta(F,F')$.

Proof. Since $\delta(F,F')<r$, from triangle inequality and Claim 17,

$$\delta(F',F_{t_1})\le\delta(F',F)+\delta(F,F_{t_1})\le r+4r\le 5r.$$

Similarly, $\delta(F',F_{t_2})\le 5r$. Then from Claim 18,

$$\ell^{\mathsf{con}}(\phi(F'),t_1)\le 2\cdot(\delta(F',F_{t_1})-\nu(F'))\le 2\cdot\delta(F',F_{t_1})\le 10r,$$

and symmetrically, $\ell^{\mathsf{con}}(\phi(F'),t_2)<10r$. Therefore,

$$\ell^{\mathsf{con}}(\phi(F'),t_1)+\ell^{\mathsf{con}}(\phi(F'),t_2)+\ell^{\mathsf{con}}(t_1,t_2)<20r+4r<g.$$

Therefore, the point $\phi(F)$ must lie either on the $\phi(F')$-t_1 shortest path or the $\phi(F')$-t_2 shortest path in V^{con}, as otherwise the union of $\phi(F')$-t_1 shortest path, $\phi(F')$-t_2 shortest path, and t_1-t_2 shortest path in V^{con} contains a cycle of length less than g, a contradiction.

Assume without loss of generality that $\phi(F)$ lies on the $\phi(F')$-t_1 shortest path. Then

$$\begin{aligned}\ell^{\mathsf{con}}(\phi(F),\phi(F')) &= \ell^{\mathsf{con}}(\phi(F'),t_1(F))-\ell^{\mathsf{con}}(\phi(F),t_1(F))\\ &\le 2\Big(\delta(F',t_1(F))-\nu(F')\Big)-2\Big(\delta(F,t_1(F))-\nu(F)\Big)\\ &\le 2\Big(\delta(F,F')+\nu(F)-\nu(F')\Big).\end{aligned}$$

Similarly, $\ell^{\mathsf{con}}(\phi(F),\phi(F'))\le 2\big(\delta(F,F')+\nu(F')-\nu(F)\big)$. Altogether, $\ell^{\mathsf{con}}(\phi(F),\phi(F'))\le 2\delta(F,F')$. ◁

We now complete the construction of the mapping ϕ, by specifying the images of all clusters $F \in \mathcal{F}$ with $A_F > r$. Let t^* be an arbitrarily chosen terminal in T. For all clusters $F \in \mathcal{F}$ with $A_F > r$, we simply set $\phi(F) = t^*$.

It remains to show that the second condition in Lemma 14 holds for all pairs $F, F' \in \mathcal{F}$. Consider a pair F, F'. Assume first that $\delta(F, F') > g/60$. Then

$$\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le \mathsf{diam}(G) \le \frac{60 \cdot \mathsf{diam}(G)}{\mathsf{girth}(G)} \cdot \delta(F, F').$$

Assume now that $\delta(F, F') \le g/60$, and without loss of generality that $A_F \le A_{F'}$. Note that, from triangle inequality, for every terminal $t \in T$, $\delta(F', F_t) \le \delta(F, F_t) - \delta(F, F')$. This implies that $A_{F'} - A_F \le \delta(F, F')$. Therefore, the probability that the random number r takes value from the interval $[A_F, A_{F'}]$ is at most $\frac{\delta(F,F')}{g/60}$. Note that, if $r \le A_F$, then $\phi(F) = \phi(F') = t^*$ and $\ell^{\mathsf{con}}(\phi(F), \phi(F')) = 0$. And if $r \ge A_{F'}$, then from Claim 19, $\ell^{\mathsf{con}}(\phi(F), \phi(F')) \le 2\delta(F, F')$. Altogether,

$$\mathbb{E}\left[\ell^{\mathsf{con}}(\phi(F), \phi(F'))\right] \le \frac{60 \cdot \mathsf{diam}(G)}{\mathsf{girth}(G)} \cdot \delta(F, F') + 2\delta(F, F') \le O\left(\frac{\mathsf{diam}(G)}{\mathsf{girth}(G)}\right) \cdot \delta(F, F').$$

References

1 Alexandr Andoni, Anupam Gupta, and Robert Krauthgamer. Towards (1+ε)-approximate flow sparsifiers. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 279–293. SIAM, 2014.

2 Haris Angelidakis, Yury Makarychev, and Pasin Manurangsi. An improved integrality gap for the călinescu-karloff-rabani relaxation for multiway cut. In *Integer Programming and Combinatorial Optimization: 19th International Conference, IPCO 2017, Waterloo, ON, Canada, June 26-28, 2017, Proceedings*, pages 39–50. Springer, 2017.

3 Aaron Archer, Jittat Fakcharoenphol, Chris Harrelson, Robert Krauthgamer, Kunal Talwar, and Éva Tardos. Approximate classification via earthmover metrics. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1079–1087. Society for Industrial and Applied Mathematics, 2004.

4 Kristóf Bérczi, Karthekeyan Chandrasekaran, Tamás Király, and Vivek Madan. Improving the integrality gap for multiway cut. *Mathematical Programming*, 183(1-2):171–193, 2020.

5 Niv Buchbinder, Joseph Naor, and Roy Schwartz. Simplex partitioning via exponential clocks and the multiway cut problem. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 535–544, 2013.

6 Niv Buchbinder, Roy Schwartz, and Baruch Weizman. Simplex transformations and the multiway cut problem. In *Proceedings of the twenty-eighth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2400–2410. SIAM, 2017.

7 Gruia Călinescu, Howard Karloff, and Yuval Rabani. An improved approximation algorithm for multiway cut. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 48–52, 1998.

8 Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing*, 34(2):358–372, 2005.

9 Moses Charikar, Tom Leighton, Shi Li, and Ankur Moitra. Vertex sparsifiers and abstract rounding algorithms. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 265–274. IEEE, 2010.

10 Chandra Chekuri, Sanjeev Khanna, Joseph Naor, and Leonid Zosin. A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM Journal on Discrete Mathematics*, 18(3):608–625, 2004.

11 Yu Chen and Zihan Tan. Towards the characterization of terminal cut functions: a condition for laminar families. *arXiv preprint*, 2023. `arXiv:2310.11367`.

12 Yu Chen and Zihan Tan. On (1+ ε)-approximate flow sparsifiers. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1568–1605. SIAM, 2024.

13 Julia Chuzhoy. On vertex sparsifiers with steiner nodes. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 673–688, 2012.

14 Julia Chuzhoy and Joseph Naor. The hardness of metric labeling. *SIAM Journal on Computing*, 36(5):1376–1386, 2007.

15 Elias Dahlhaus, David S. Johnson, Christos H. Papadimitriou, Paul D. Seymour, and Mihalis Yannakakis. The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23(4):864–894, 1994.

16 Jittat Fakcharoenphol, Chris Harrelson, Satish Rao, and Kunal Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 257–265. Society for Industrial and Applied Mathematics, 2003.

17 Ari Freund and Howard Karloff. A lower bound of $8/(7+\frac{1}{k-1})$ on the integrality ratio of the călinescu–karloff–rabani relaxation for multiway cut. *Information Processing Letters*, 75(1-2):43–50, 2000.

18 Alan Frieze. Hamilton cycles in the union of random permutations. *Random Structures & Algorithms*, 18(1):83–94, 2001.

19 Torben Hagerup, Jyrki Katajainen, Naomi Nishimura, and Prabhakar Ragde. Characterizing multiterminal flow networks and computing flows in networks of small treewidth. *Journal of Computer and System Sciences*, 57(3):366–375, 1998.

20 Howard Karloff, Subhash Khot, Aranyak Mehta, and Yuval Rabani. On earthmover distance, metric labeling, and 0-extension. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 547–556, 2006.

21 Howard Karloff, Subhash Khot, Aranyak Mehta, and Yuval Rabani. On earthmover distance, metric labeling, and 0-extension. *SIAM Journal on Computing*, 39(2):371–387, 2009.

22 Alexander V Karzanov. Minimum 0-extensions of graph metrics. *European Journal of Combinatorics*, 19(1):71–101, 1998.

23 Jon Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. *Journal of the ACM (JACM)*, 49(5):616–639, 2002.

24 Robert Krauthgamer and Ron Mosenzon. Exact flow sparsification requires unbounded size. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2354–2367. SIAM, 2023.

25 William Kuszmaul and Qi Qi. The multiplicative version of azuma's inequality, with an application to contention analysis. *arXiv preprint*, 2021. `arXiv:2102.05077`.

26 F Thomson Leighton and Ankur Moitra. Extensions and limits to vertex sparsification. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 47–56. ACM, 2010.

27 Konstantin Makarychev and Yury Makarychev. Metric extension operators, vertex sparsifiers and lipschitz extendability. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 255–264. IEEE, 2010.

28 Rajsekar Manokaran, Joseph Seffi Naor, Prasad Raghavendra, and Roy Schwartz. Sdp gaps and ugc hardness for multiway cut, 0-extension, and metric labeling. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 11–20. ACM, 2008.

29 Ankur Moitra. Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 3–12. IEEE, 2009.

30 Doron Puder. Expansion of random graphs: New proofs, new results. *Inventiones mathematicae*, 201(3):845–908, 2015.

31 Roy Schwartz and Nitzan Tur. The metric relaxation for 0-extension admits an $\omega(\log^{2/3} k)$ gap. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1601–1614, 2021.

A Comparison between 0EwSN and the variant in [1]

The Steiner node variant of 0-Ext in [1]

In [1], the following problem (referred to as $\mathsf{0EwSN_{AGK}}$) was proposed.

The input consists of

- an edge-capacitated graph $G = (V, E, c)$, with length $\{\ell_e\}_{e \in E}$ on its edges;
- a set $T \subseteq V$ of k terminals; and
- a *demand* $\mathcal{D} : T \times T \to \mathbb{R}^+$ on terminals.

A solution consists of

- a partition $\mathcal{F}$ of V with $|\mathcal{F}|$, such that distinct terminals of T belong to different sets in $\mathcal{F}$; for each vertex $u \in V$, we denote by $F(u)$ the cluster in $\mathcal{F}$ that contains it;
- a semi-metric δ on the clusters in $\mathcal{F}$, such that: $\sum_{t,t'} \mathcal{D}(t,t') \cdot \delta(F(t), F(t')) \geq \sum_{t,t'} \mathcal{D}(t,t') \cdot \mathsf{dist}_\ell(t,t')$, where $\mathsf{dist}_\ell(\cdot,\cdot)$ is the shortest-path distance (in G) metric induced by edge length $\{\ell_e\}_{e \in E(G)}$.

The cost of a solution $(\mathcal{F}, \delta)$ is $\mathsf{cost}(\mathcal{F}, \delta) = \sum_{(u,v) \in E} c(u,v) \cdot \delta(F(u), F(v))$, and its *size* is $|\mathcal{F}|$.

From (LP1) and Proposition 4.2 in [1], it is proved that:

▶ **Proposition 20.** *Given a graph G and a subset T of its terminals, and a function f, if for every length $\{\ell_e\}_{e \in E(G)}$ and every demand D, the instance $(G, T, \ell, \mathcal{D})$ of $\mathsf{0EwSN}_{AGK}$ has a solution $(\mathcal{F}, \delta)$ with size $|\mathcal{F}| \leq f(k)$ and cost*

$$\sum_{(u,v) \in E} c(u,v) \cdot \delta(F(u), F(v)) \leq q \cdot \sum_{(u,v) \in E} c(u,v) \cdot \mathsf{dist}_\ell(u,v),$$

then there is a quality-$(1+\varepsilon)q$ flow sparsifier H for G w.r.t T with $|V(H)| \leq (f(k))^{(\log k/\varepsilon)^{k^2}}$.

The 0EwSN problem

In studying the integrality gap of (LP-Metric), we are essentially considering the following problem.

The input consists of

- an edge-capacitated graph $G = (V, E, c)$, with length $\{\ell_e\}_{e \in E}$ on its edges; and
- a set $T \subseteq V$ of k terminals.

A solution consists of

- a partition $\mathcal{F}$ of V with $|\mathcal{F}|$, such that distinct terminals of T belong to different sets in $\mathcal{F}$; for each vertex $u \in V$, we denote by $F(u)$ the cluster in $\mathcal{F}$ that contains it;
- a semi-metric δ on the clusters in $\mathcal{F}$, such that for all pairs $t, t' \in T$, $\delta(F(t), F(t')) = \mathsf{dist}_\ell(t,t')$, where $\mathsf{dist}_\ell(\cdot,\cdot)$ is the shortest-path distance (in G) metric induced by edge length $\{\ell_e\}_{e \in E(G)}$.

The cost and the size of a solution is defined in the same way as $\mathsf{0EwSN_{AGK}}$.

The difference between two problems are underlined. Specifically, in $\mathsf{0EwSN_{AGK}}$ it is only required that some "average terminal distance" does not decrease, while in our problem it is required that all pairwise distances between terminals are preserved. Clearly, our requirement for a solution is stronger, which implies that any valid solution to our instance is also a valid solution to the same $\mathsf{0EwSN_{AGK}}$ instance (with arbitrary $\mathcal{D}$). Therefore, we have the following corollary.

▶ **Corollary 21.** *Given a graph G and a subset T of its terminals, and a function f, if for every length $\{\ell_e\}_{e\in E(G)}$, the instance (G,T,ℓ) of* 0EwSN *has a solution $(\mathcal{F},\delta)$ with size $|\mathcal{F}|\le f(k)$ and cost*

$$\sum_{(u,v)\in E} c(u,v)\cdot\delta(F(u),F(v)) \le q\cdot \sum_{(u,v)\in E} c(u,v)\cdot \mathsf{dist}_\ell(u,v),$$

then there is a quality-$(1+\varepsilon)q$ flow sparsifier H for G w.r.t T with $|V(H)|\le (f(k))^{(\log k/\varepsilon)^{k^2}}$.

On the other hand, the main result of our paper is a lower bound for the 0EwSN problem. As 0EwSN has stronger requirement (for solutions) than $\mathsf{0EwSN}_{\mathsf{AGK}}$, our lower bound does not immediately imply a lower bound for $\mathsf{0EwSN}_{\mathsf{AGK}}$ or for the flow sparsifier. However, if we can show that, for some function f, there exists a graph G, a terminal set T with size k, and a demand $\mathcal{D}$ on T, such that any solution $(\mathcal{F},\delta)$ with size $|\mathcal{F}|\le f(k)$ has cost at least

$$\sum_{(u,v)\in E} c(u,v)\cdot\delta(F(u),F(v)) \ge q\cdot \sum_{(u,v)\in E} c(u,v)\cdot \mathsf{dist}_\ell(u,v),$$

then this, from the (LP1) and the discussion in [1], implies that any quality-$o(q)$ contraction-based flow sparsifier for G has size at least $f(k)$.