RC22578 (W0210-022) October 3, 2002
Mathematics

IBM Research Report

Sparse Distance Preservers and Additive Spanners

[2]Bela Bollobas, [1]Don Coppersmith, [3]Michael Elkin

[1]IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

[2]Microsoft Research,
One Microsoft Way,
Redmon ,WA 98122

[3]School of Mathematics,
Institute for Advanced Study,
Princeton, NJ,08540.

Sparse Distance Preservers and Additive Spanners

Béla Bollobás [*†‡§] Don Coppersmith [¶] Michael Elkin [‖**]

Abstract

For an unweighted graph $G = (V, E)$, $G' = (V, E')$ is a subgraph if $E' \subseteq E$, and $G'' = (V'', E'', \omega)$ is a *Steiner graph* if $V \subseteq V''$, and for any pair of vertices $u, w \in V$, the distance between them in G'' (denoted $d_{G''}(u, w)$) is at least the distance between them in G (denoted $d_G(u, w)$).

In this paper we introduce the notion of *distance preserver*. A subgraph (resp., Steiner graph) G' of a graph G is a subgraph (resp., Steiner) *D-preserver* of G if for every pair of vertices $u, w \in V$ with $d_G(u, w) \geq D$, $d_{G'}(u, w) = d_G(u, w)$. We show that any graph (resp., digraph) has a *subgraph* D-preserver with at most $O(n^2/D)$ edges (resp., arcs), and there are graphs and digraphs for which any *undirected Steiner* D-preserver contains $\Omega(n^2/D)$ edges. However, we show that if one allows a *directed Steiner* (or, shortly, *diSteiner*) D-preserver, then these bounds can be improved. Specifically, we show that for any graph or digraph there exists a diSteiner D-preserver with $O(\frac{n^2 \cdot \log D}{D \cdot \log n})$ arcs, and that this result is tight up to a constant factor.

We also study *D-preserving* distance labeling schemes, that are labeling schemes that guarantee precise calculation of distances between pairs of vertices that are at distance at least D one from another. We show that there exists a D-preserving labeling scheme with labels of size $O(\frac{n}{D} \log^2 n)$, and that labels of size $\Omega(\frac{n}{D} \log D)$ are required for any D-preserving labeling scheme.

Finally, we study *additive* spanners. A subgraph G' of an undirected graph $G = (V, E)$ is its additive β-spanner if for any pair of vertices $u, w \in V$, $d_{G'}(u, w) \leq d_G(u, w) + \beta$. It is known that for any n-vertex graph there exists an additive 2-spanner with $O(n^{3/2})$ edges, and an additive Steiner 4-spanner with $O(n^{4/3})$ edges. However, no construction of additive spanners with $o(n^{3/2})$ edges or Steiner additive spanners with $o(n^{4/3})$ edges are known so far. We devise a construction of additive $O(2^{1/\delta} n^{(1-\delta)\frac{\lceil 1/\delta \rceil - 2}{\lceil 1/\delta \rceil - 1}})$-spanner with $O(n^{1+\delta})$ edges for any graph and any $\delta > 0$[1].

[*]Microsoft Research, One Microsoft Way, Redmond, WA 98122

[†]Department of Mathematical Sciences, University of Memphis, Memphis TN 38152

[‡]Trinity College, Cambridge CB2 1TQ, UK

[§]Research supported by NSF grant DSM 9971788 and DARPA grant F33615-01-C-1900

[¶]IBM Research, Yorktown Heights, NY, USA; dcopper@us.ibm.com

[‖]School of Mathematics, Institute for Advanced Study, Princeton, NJ, USA, 08540. elkin@ias.edu. This material is based upon work supported by the National Science Foundation under agreement No. DMS-9729992. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[**]Part of this work was done in Weizmann Institute, Department of Computer Science and Mathematics, Rehovot, Israel.

[1]Preliminary version of this paper appeared in SODA 2003 [7].

1 Introduction

A graph $G' = (V, E')$ is a *subgraph* of an unweighted graph $G = (V, E)$ if $E' \subseteq E$. The *distance* from a vertex u to a vertex w in G, denoted $d_G(u, w)$, is the number of edges in the shortest (in terms of the number of edges) path from u to w in G. Note that the distances in a subgraph G' may be only greater than the corresponding distances in G. A (possibly weighted) graph $G' = (V', E', \omega)$ is a *Steiner graph* of G if $V \subseteq V'$, and for any pair of vertices $u, w \in V$, $d_{G'}(u, w) \geq d_G(u, w)$, and for any edge $e' \in E'$, $\omega(e') \geq 0$. Observe that any subgraph G' of G is, in particular, its Steiner graph. A subgraph or a Steiner graph G' of G that *approximates* (in some sense) all the distances in G is called a *spanner*. In particular, for a positive integer parameter κ, G' is a κ*-spanner* of G, if for any pair of vertices u, w in G, $d_{G'}(u, w) \leq \kappa \cdot d_G(u, w)$. The number κ is called the *stretch* or *distortion* factor of the spanner G'.

Spanners were intensively studied during the last fifteen years. They have multiple applications in distributed computing [2, 3, 22, 14, 4] and computational geometry [9, 12]. Furthermore, constructing a spanner and applying existing algorithms on it was used as an algorithmic technique in [4, 10, 11, 14].

Peleg and Schäffer [21] have shown that for any positive integer κ and any n-vertex graph G there exists a subgraph $O(\kappa)$-spanner G' with $O(n^{1+1/\kappa})$ edges. Note that this result indicates a *tradeoff* between the stretch of the spanner and the number of edges it uses. This tradeoff was shown to be essentially the best possible in [21], but some constant factors were improved later on in [1, 8]. These papers also generalized the result to *weighted* graphs. Recently, Elkin and Peleg [15, 14] have shown that the aforementioned tradeoff is tight only as far as the distortion of *small distances* is considered, and can be almost eliminated whenever one is interested in approximating the distances that are greater than certain constant. Specifically, it is shown there that for any pair of parameters $\epsilon > 0$, $\kappa = 1, 2, \ldots$ there exists a threshold $\beta = \beta(\epsilon, \kappa)$ such that for any n-vertex graph G there exists a subgraph spanner G' with $O(n^{1+1/\kappa})$ edges such that for any pair of vertices u, w that are at distance at least β one from another in G, the distance between in G' is at most by a factor $1 + \epsilon$ greater than the one in G (i.e., $d_{G'}(u, w) \leq (1 + \epsilon) \cdot d_G(u, w)$). In other words, *large distances* can be approximated arbitrarily well by an arbitrarily sparse spanners. In view of this result due to [15], it is natural to ask whether approximation is at all necessary whenever large distances are under consideration, or, maybe large distances can be *preserved* using a sparse spanner.

To address this question, we introduce a notion of a *distance preserving subgraph*, briefly, a *preserver*. A subgraph G' of a graph G is a D*-preserver* of G if for every pair of vertices $u, w \in V$ with $d_G(u, w) \geq D$, $d_{G'}(u, w) = d_G(u, w)$. (The same definition applies to Steiner graphs as well.) We show that any graph (respectively, digraph) has a *subgraph* D-preserver with at most $O(n^2/D)$ edges (resp., arcs), and there are graphs and digraphs for which any *undirected Steiner* D-preserver contains $\Omega(n^2/D)$ edges (resp., arcs). However, we show that if one allows a *directed Steiner* (or, shortly, *diSteiner*) D-preserver, then these bounds can be improved. Specifically, we show that for any graph or digraph there exists a diSteiner D-preserver with $O(\frac{n^2 \cdot \log D}{D \cdot \log n})$ arcs, and that this result is tight up to a constant factor. In particular, it follows that for any graph or digraph there is a diSteiner 1-preserver with $O(n^2/\log n)$ arcs. Generalizing this result, we show that for any graph (resp., digraph) with $m \geq c' \cdot n^{3/2}$ edges (resp., arcs), for some small constant $c' > 1$, there is a diSteiner 1-preserver with fewer than m arcs, and that a factor of $\frac{\log n}{c \log \log n}$ (resp., $\log^{1-\gamma} n$) can be "saved" for $m = n^2/\log^c n$ (resp., $m = n^2/2^{\log^\gamma n}$) for any $c > 0$ (resp., $0 < \gamma < 1$). We also show that for any bipartite graph with m edges and girth greater than 4, any diSteiner 1-preserver contains at least m arcs, and as there are such graphs with $m = (1/2 + o(1))n^{3/2}$ edges, it follows that this upper bound cannot be generalized to graphs with $m \leq (1/2)n^{3/2}$ edges.

Our proof of the existence of sparse diSteiner preservers uses the following theorem.

Theorem 1.1 *(cf. [5]) Let G be an n-vertex graph with average degree d, and $t = 1, 2, \ldots$, $s = t, t+1, \ldots$, such that $n\binom{d}{t} > (s-1)\binom{n}{t}$. Then G contains a $K_{s,t}$ (complete bipartite subgraph with one bipartition of size s and another of size t).*

In order to convert our proof of *existence* of diSteiner D-preservers into a polynomial time algorithm for computing them, we devised a constructive proof of Theorem 1.1. This proof might be of independent interest in the context of Ramsey theory. From algorithmic perspective, this proof may serve as an algorithm for computing a subgraph isomorphic to $K_{s,t}$ in a graph that satisfies the assumptions of Theorem 1.1. The complexity of this algorithm is $O(n^2 \cdot t)$. We use this result for devising an algorithm with running time $O(n^4 \frac{(\log\log n)^2}{\log n})$ (resp., $O(m^3 \cdot n)$) for computing a diSteiner 1-preserver (resp., D-preserver) with $O(n^2/\log n)$ (resp., $O(\frac{n^2 \log D}{D \cdot \log n})$) arcs for an arbitrary n-vertex graph with m edges. We remark that any improvement of a factor of $\Omega(n)$ in the running time of an algorithm for constructing a diSteiner 1-preserver would have some interesting applications to efficient computation of distances in dense graphs (by computing their diSteiner 1-preserver, and performing distance computations on the 1-preserver, assuming that the latter is sparser than the original graph).

In particular, our results address the aforementioned question and show that *approximation* of large distances is indeed necessary as far as arbitrarily sparse spanners are considered, as there exist infinite families of graphs in which large distances cannot be preserved by a *sparse* spanner.

We also generalize the definition of D-preserver, and say that G' is a (D, g)*-preserver* of G if for any pair of vertices $u, w \in V$ such that $d_G(u, w) \geq D$, we have $d_{G'}(u, w) \leq d_G(u, w) + g$. In this context, we show upper and lower bounds on the maximal number m_1 of edges in a graph for which any subgraph (D, g)-preserver contains at least m_1 edges. We show that $\Omega(\frac{n^{1+c_0/(g+2)}}{g \cdot D^{c_0/(g+2)}}) = m_1 = O(\frac{n^{1+1/\lfloor g/4 \rfloor}}{D^{1/\lfloor g/4 \rfloor}})$, where $4/3 \leq c_0 \leq 2$, and under Erdős girth conjecture, $c_0 = 2$. The lower bound serves also as a lower bound on the minimal number m_2 such that any graph has a subgraph (D, g)-preserver with m_2 edges. However, so far we were not able to prove a non-trivial upper bound on the size of (D, g)-preservers, and, in particular, it is not clear to us whether these two dual notions m_1 and m_2 are equal.

We also study the problem of preserving long distances in the context of *distance labeling schemes*. Distance labeling scheme is a pair of functions $(\mathcal{M}, \mathcal{D})$. The *labeling* function $\mathcal{M}$, given a graph G and a vertex v, returns a bit string, often called the *label of v*. The *query-answering* function $\mathcal{D}$, given a pair of labels, returns an estimate of the distance between the corresponding pair of vertices.

The problem of devising distance labeling schemes with *short* labels was introduced in [20], and is intensively studied till then [17, 25, 24]. We consider *D-preserving* labeling schemes, that are schemes that satisfy $\mathcal{D}(\mathcal{M}(G, u), \mathcal{M}(G, w)) = d_G(u, w)$ for any graph $G = (V, E)$ and pair of vertices $u, w \in V$ such that $d_G(u, w) \geq D$. We show that there exists a D-preserving labeling scheme with labels of size $O(\frac{n}{D} \log^2 n)$, and that labels of size $\Omega(\frac{n}{D} \log D)$ are required for any D-preserving labeling scheme.

Finally, we study *additive* spanners. A subgraph G' of an undirected graph $G = (V, E)$ is its additive β-spanner if for any pair of vertices $u, w \in V$, $d_{G'}(u, w) \le d_G(u, w) + \beta$. (The same definition applies to Steiner graphs as well.) It is known [13, 15] that for any unweighted undirected n-vertex graph there exists an additive 2-spanner with $O(n^{3/2})$ edges, and an additive Steiner 4-spanner with $O(n^{4/3})$ edges. However, to the best of our knowledge, no construction of additive spanners with $o(n^{3/2})$ edges or Steiner additive spanners with $o(n^{4/3})$ edges are known.

It is implicit in [15] that the existence of a D-preserver with m edges for a graph implies the existence of an additive $O(D)$-spanner with the same number of edges for the same graph. Hence, our aforementioned results concerning D-preservers imply that for any n-vertex graph and for any $\delta > 0$ there exists an additive $O(n^{1-\delta})$-spanner with $O(n^{1+\delta})$ edges. We improve upon this and devise a construction of additive $O(2^{1/\delta} n^{(1-\delta)\frac{\lceil 1/\delta \rceil - 2}{\lceil 1/\delta \rceil - 1}})$-spanners with $O(n^{1+\delta})$ edges for any graph and any $\delta > 0$. In particular, this implies a construction of additive $O(n^{1/4+\epsilon/2})$-spanners with $O(n^{3/2-\epsilon})$ edges and $O(n^{1/3})$-spanners with $O(n^{4/3})$ edges and $O(n^{4/9+(2/3)\epsilon})$-spanner with $O(n^{1/3-\epsilon})$ edges (for additive Steiner spanners we achieve slightly better results). This construction is based on the construction of $(1+\epsilon, \beta)$-spanners due to Elkin and Peleg [15], and, in addition, uses distance preservers. We provide a short sketch of this construction in this paper; the complete proof will be described elsewhere.

Related work: After our basic results (the existence of subgraph D-preserver with $O(n^2/D)$ edges and the lower bound of $\Omega(n^2/D)$ on the number of edges in subgraph D-preservers) were communicated to Mikkel Thorup, he devised [23] a more efficient randomized procedure for computing a subgraph D-preserver of size $O(n^2 \log n/D)$ (greater than optimal by a logarithmic factor). This more efficient procedure uses some techniques of [26] from the area of dynamic algorithms. The efficiency of the procedure of [23] makes it more suitable for algorithmic applications such as (and this is, indeed, the motivation of [23]) computing shortest paths between pairs of vertices that are at distance at least D one from another. We use a similar idea to devise D-preserving labeling schemes.

Our algorithm for constructing sparse diSteiner 1-preservers for general graphs successively extracts large bipartite cliques and replaces them by directed stars. Similar idea of extracting large bipartite cliques was used by Feder and Motwani in [16] for constructing *compressions* of graphs. The notion of compression graph is somewhat similar to the notion of Steiner graph, but the distances in compression graph may be *shorter* than the distances in the original graph.

Structure of the paper: In Section 3 we show some preliminary results concerning additive spanners that are derived quite easily from [15]. In Section 4 we discuss the issue of distance preservation, which is the main topic of this paper. This section is divided into Subsection 4.2, that is devoted to the lower bounds, and Subsection 4.3, that is devoted to the upper bounds. In Section 4.3.2 we address the algorithmic aspects of our paper. In particular, this section contains our constructive proof of Theorem 1.1 and a description of a D-preserving labeling scheme. Finally, in Section 5 we sketch the proof of our results concerning the additive spanners.

2 Preliminaries

Given a digraph (resp., undirected graph) $G = (V, E)$, a sequence of vertices $P = (v_0, v_1, \ldots, v_s)$, $s \ge 0$, is called a *walk* if $\langle v_i, v_{i+1} \rangle$ (resp., (v_i, v_{i+1})) belongs to E, for any $i = 0, 1, \ldots, s-1$. A walk $P = (v_0, v_1, \ldots, v_s)$ is a *path*, if $v_i \neq v_j$ for any $i, j = 0, 1, \ldots, s$, $i \neq j$.

The *head* (resp., *tail*) of P, denoted $head(P)$ (resp., $tail(P)$) is v_0 (resp., v_s). Given a path $P = (v_0, v_1, \ldots, v_s)$, and an arc $\langle v_i, v_{i+1} \rangle \in P$, let $prefix(P, \langle v_i, v_{i+1} \rangle)$ (resp., $suffix(P, \langle v_i, v_{i+1} \rangle)$) denote the path $(v_0, \ldots, v_i)$ (resp., $(v_{i+1}, \ldots, v_s)$).

For a digraph (resp., undirected graph) $G = (V, E)$, and an arc (resp., edge) $e \in E$, let G_e denote the digraph (resp., undirected graph) $(V, E \setminus \{e\})$.

In an undirected graph $G = (V, E)$, given a walk $P = (v_0, \ldots, v_s)$, and an edge $e = (v_i, v_{i+1}) \in P$, the (e, v_i)-*endpoint* of P, denoted $endpoint(P, e, v_i)$, is v_0. The (e, v_i)-*subpath* of P, denoted by $subpath(P, e, v_i)$, is $(v_0, \ldots, v_i)$.

Given two walks $P_1 = (v_0, \ldots, v_s)$ and $P_2 = (v_s, \ldots, v_{t+s})$, $t, s \geq 0$, the *concatenation* $P_1 \cdot P_2$ is the walk $(v_0, \ldots, v_{t+s})$. Obviously, the concatenation is associative, and so $P_1 \cdot P_2 \cdot \ldots \cdot P_r$ is well-defined, whenever for any $i = 1, \ldots, r-1$, $P_i \cdot P_{i+1}$ is defined.

Given a (directed or undirected) graph $G = (V, E)$, and a pair of vertices $u, w \in V$, let the *distance between u and w in G*, denoted $d_G(u, w)$ or $d_E(u, w)$, be the length of the shortest path from u to w in G. If no such a path exists, the distance is defined to be equal to infinity.

Let $G = (V, E)$ be a (directed or undirected) graph, and $v \in V$ be a vertex. Let $Out(v, G)$ (resp., $In(v, G)$) denote the set $\{u \in V \mid d_G(v, u) \neq \infty\}$ (resp., $\{u \in V \mid d_G(u, v) \neq \infty\}$).

A digraph $T = (V, E_T)$ is an *out-tree* (resp., *in-tree*) if it is acyclic and connected in the undirected sense and there is a vertex $v \in V$, called the *root*, such that for any $w \in V$ there exists a unique directed path in T from v to w (resp., from w to v).

Given a digraph $G = (V, E)$, and a vertex $v \in V$, an *out-tree* (resp., *in-tree*) $T = (V', E_T)$ rooted at v is called the *BFS spanning out-tree* (resp., *in-tree*) of G rooted at v, denoted $T_{out}(v, G)$ (resp., $T_{in}(v, G)$), if $V' = Out(v, G)$ (resp., $V' = In(v, G)$), and for any vertex $w \in Out(v, G)$ (resp., $w \in In(v, G)$), $d_T(v, w) = d_G(v, w)$ (resp., $d_T(w, v) = d_G(w, v)$).

In an undirected graph $G = (V, E)$, a sequence of vertices $C = (v_0, v_1, \ldots, v_s, v_0)$ is a *cycle*, if $v_i \in V$ for any $i = 0, 1, \ldots, s$, $v_i \in V$, and for any $i = 0, 1, \ldots, s-1$, $(v_i, v_{i+1}) \in E$ and $(v_s, v_0) \in E$. The *length* of the cycle C is $s + 1$.

For a graph $G = (V, E)$, a vertex $v \in V$, and integer $k = 0, 1, 2, \ldots$, let $\Gamma_k(v, G)$ (resp., $\hat{\Gamma}_k(v, G)$) denote the set of vertices that are at distance precisely (resp., at most) k from v, i.e., $\Gamma_k(v, G) = \{u \in V \mid d_G(v, u) = k\}$, $\hat{\Gamma}_k(v, G) = \{u \in V \mid d_G(v, u) \leq k\}$.

Given a digraph $G = (V, E)$, and a positive integer distance threshold D, the *D-path associated with an arc e*, denoted by $P(e, D)$, is one of the shortest paths between its endpoints $head(P(e, D))$ and $tail(P(e, D))$ such that

$$d_G(head(P(e, D)), tail(P(e, D))) = |P(e, D)| \geq D \ , \tag{1}$$

$$d_{G_e}(head(P(e, D)), tail(P(e, D))) > d_G(head(P(e, D)), tail(P(e, D))) \ . \tag{2}$$

Given an undirected graph $G = (V, E)$, and a positive integer D, the *D-path associated with the edge e*, denoted $P(e, D)$, is one of the shortest paths between $endpoint(P(e), e, v)$ and $endpoint(P(e), e, z)$ such that

$$d_G(endpoint(P(e), e, v), endpoint(P(e), e, z)) = |P(e)| \geq D \ , \tag{3}$$

$$d_{G_e}(endpoint(P(e), e, v), endpoint(P(e), e, z)) > d_G(endpoint(P(e), e, v), endpoint(P(e), e, z)) \tag{4}$$

Note that such a path may not exist, and, on the other hand, there may be several such paths. In the latter case, set $P(e, D)$ to be an arbitrary such a path.

Throughout the paper, whenever the value of D is clear from the context, we use the notation $P(e)$ instead of $P(e, D)$.

3 Additive Spanners: Preliminary Results

A subgraph G' of a graph $G = (V, E)$ is its (α, β)-*spanner* if for any pair of vertices $u, w \in V$, $d_{G'}(u, w) \leq \alpha \cdot d_G(u, w) + \beta$. Our starting point is the following result from [15].

Theorem 3.1 *[15] Given constants $0 < \epsilon, \delta < 1$, there is a constant $\beta = \beta(\delta, \epsilon) = (1/\delta)^{\max\{(\log\log 1/\delta - \log \epsilon)(1-1/\log 1/\delta), 3\}}$ such that for any graph G, there exists a constructible in polynomial time $(1+\epsilon, \beta)$-spanner $G' = (V, E')$ and Steiner $(1+\epsilon, \beta)$-spanner $G'' = (V'', E'', \omega)$ with $|E'| = O(\beta n^{1+\delta})$ and $|E''| = O(n^{1+\delta})$.*

(The result about Steiner spanners is implicit in [15].) The next lemma follows from the definitions.

Lemma 3.2 *Let G' be a (possibly Steiner) (α, β)-spanner of a graph G, and let $u, w \in V(G)$ be a pair of vertices. Then $d_{G'}(u, w) \leq d_G(u, w) + ((\alpha - 1)d_G(u, w) + \beta)$.*

Proof: By definition of (α, β)-spanner. $d_{G'}(u, w) \leq \alpha \cdot d_G(u, w) + \beta = d_G(u, w) + ((\alpha - 1)d_G(u, w) + \beta)$. ∎

Corollary 3.3 *An (α, β)-spanner G' of an n-vertex graph G is an additive $((\alpha - 1) \cdot n + \beta)$-spanner of G.*

Obviously, the same statement is true for Steiner spanners as well. Theorem 3.1 and Lemma 3.2 imply

Lemma 3.4 *Given $n = 2, 3, \ldots$, $\Omega(1/\log n) = \delta < 1$, $t = 1, 2, \ldots, n-1$ and an n-vertex graph $G = (V, E)$, there exists a subgraph $G' = (V, E')$, $|E'| = O(n^{1+\delta} t^{\delta})$, and Steiner graph $G'' = (V'', E'', \omega)$, $|E''| = O(n^{1+\delta})$, such that for any pair of vertices $u, w \in V$ such that $d_G(u, w) \leq t$,*

$$\begin{aligned} d_{G'}(u, w) &\leq d_G(u, w) \\ &+ O(1/\delta \cdot t^{1 - \frac{1}{(1/\delta)(\log(1/\delta) - 1)}}) \ , \end{aligned} \tag{5}$$

$$\begin{aligned} d_{G''}(u, w) &\leq d_G(u, w) \\ &+ O((t \cdot \log(1/\delta))^{1 - 1/\log(1/\delta)}) \ . \end{aligned} \tag{6}$$

Proof: By Theorem 3.1, for any $\epsilon, \delta > 0$, and for any n-vertex graph G, there exists a Steiner $(1+\epsilon, \beta)$-spanner, $\beta = \beta(\delta, \epsilon)$ with $O(n^{1+\delta})$ edges. By Lemma 3.2, for any pair of vertices $u, w \in V$,

$$d_{G''}(u, w) \leq d_G(u, w) + (\epsilon \cdot d_G(u, w) + \beta) \ .$$

Hence, for any pair of vertices $u, w \in V$ such that $d_G(u, w) \leq t$,

$$d_{G''}(u, w) \leq d_G(u, w) + (\epsilon \cdot t + \beta) \ . \tag{7}$$

Set $\epsilon = \frac{8}{t^{1/\log(1/\delta)}} \cdot (\log^{1-\frac{1}{\log(1/\delta)}} 1/\delta)$. Then $\epsilon \cdot t = 8 \cdot (t \cdot \log(1/\delta))^{1-1/\log(1/\delta)}$. Straightforward computation shows also that $\beta(\delta, \epsilon) = 8 \cdot (t \cdot \log(1/\delta))^{1-1/\log(1/\delta)}$ as well. Hence, by inequality (7), $d_{G''}(u, w) \leq d_G(u, w) + 16 \cdot (t \cdot \log(1/\delta))^{1-1/\log(1/\delta)}$.

To prove inequality (6), note that analogously to (7), it follows that there exists a subgraph $G' = (V, E')$ with $O(n^{1+\delta}\beta)$ edges such that for any pair of vertices $u, w \in V$ with $d_G(u, w) \leq t$, $d_{G'}(u, w) + (\epsilon \cdot t + \beta)$. Set $\epsilon = \frac{8/\delta}{t^{\delta/(\log(1/\delta)-1)}}$. Now, a straightforward computation implies that $\epsilon \cdot t + \beta(\delta, \epsilon) \leq (8/\delta) \cdot t^{1-\frac{\delta}{\log(1/\delta)-1}} + 8 \cdot t^{\delta}$. Also, $|E'| = O(n^{1+\delta}\beta(\delta, \epsilon)) = O(n^{1+\delta}t^{\delta})$. ∎

Setting $t = n$ implies that

Corollary 3.5 *Given $n = 2, 3, \ldots$, $\Omega(1/\log n) = \delta < 1$, and an n-vertex graph G, there exists an additive $O(1/\delta \cdot n^{1-\frac{\delta}{2\log 1/\delta}})$-spanner G' and Steiner additive $O((n \log 1/\delta)^{1-1/\log(1/\delta)})$-spanner G'' of G, both with $O(n^{1+\delta})$ edges.*

Proof: The first assertion follows from Lemma 3.4 by setting $\delta' = 2\delta$. The second assertion is an immediate consequence of Lemma 3.4. ∎

Later on (Section 5) we will show that inequality (5) can be improved to (roughly) $d_{G'}(u, w) \leq d_G(u, w) + O(t^{1-\delta})$, and, consequently, the additive error of the spanner G' in Corollary 3.5 can be improved to (roughly) $O(n^{1-2\delta})$ by appropriate modification of the arguments of [15] (note that the proof of Lemma 3.4 uses Theorem 3.1 as a blackbox).

Note also that in Corollary 3.5, n may be replaced by $Diam(G)$. The obtained statement generalizes the observation that for any graph G there is an additive $Diam(G)$-spanner G' that forms a tree.

However, already the results of Lemma 3.4 suggest that a possible direction towards improving the bounds of Corollary 3.5 could be showing that distances between *remote* pairs of vertices can be *preserved* using sparse subgraphs. We elaborate on this in the next section.

4 Distance Preservation

4.1 Discussion

Note that Theorem 3.1 implies that for any fixed $\epsilon, \delta > 0$ there exists fixed $\beta' = \beta'(\delta, \epsilon)$ such that for any undirected graph $G = (V, E)$ there exists a subgraph $G' = (V, E')$, $E' \subseteq E$ with $|E'| = O(n^{1+\delta})$ edges that *approximates within a multiplicative factor of* $1 + \epsilon$ all the distances that are already greater than β'. We start with showing that this result is optimal in the sense that $(1 + \epsilon)$-approximation is necessary, and, furthermore, for any fixed $\delta > 0$ there is no fixed $\beta' = \beta'(\delta)$ such that for any undirected graph $G = (V, E)$ there exists a subgraph $G' = (V, E')$, $E' \subseteq E$ with $|E'| = O(n^{1+\delta})$ edges that *preserves* all the distances already greater than β'.

To facilitate the discussion, let us introduce some definitions.

Definition 4.1 *For $D = 1, 2, \ldots$, a subgraph $G' = (V, E')$ of a graph $G = (V, E)$ is said to be a* (subgraph) D-preserver *of G, if for any pair of vertices $u, w \in V$ with $d_G(u, w) \geq D$, $d_{G'}(u, w) = d_G(u, w)$.*

The definition extends in a natural way to *Steiner D-preservers.*

Definition 4.2 *For $n = 2, 3, \ldots$ and $D = 1, 2, \ldots, n-1$, let $f(D, n)$ (resp., $f_S(D, n)$) be the minimal number such that for any n-vertex graph there exists a subgraph (resp., Steiner) D-preserver with at most $f(D, n)$ (resp., $f_S(D, n)$) edges. Also, let $\bar{f}(D, n)$ (resp., $\bar{f}_S(D, n)$) be the maximal number m of edges in an n-vertex graph whose any subgraph (resp., Steiner) D-preserver contains at least m edges.*

On *directed* graphs, let $f^{dir}(D, n)$, $\bar{f}^{dir}(D, n)$, $f_S^{dir}(D, n)$ and $\bar{f}_S^{dir}(D, n)$ denote the corresponding quantities.

The equality between these dual notions follows from their definitions.

Lemma 4.3 *For $n = 2, 3, \ldots$ and $D = 1, 2, \ldots, n-1$, $f(D, n) = \bar{f}(D, n)$.*

Proof: By definition of $\bar{f}(D, n)$, there exists an n-vertex graph G_0 with $\bar{f}(D, n)$ edges whose any D-preserver contains at least $\bar{f}(D, n)$ edges. By definition of $f(D, n)$, for any n-vertex graph G, there exists a D-preserver with at most $f(D, n)$ edges. In particular, there is a D-preserver of G_0 with $m' \leq f(D, n)$ edges. As $m' \geq \bar{f}(D, n)$, it follows that $\bar{f}(D, n) \leq f(D, n)$.

For the opposite direction, note that by the definition of $f(D, n)$, there exists an n-vertex graph $G_1 = (V_1, E_1)$ such that any D-preserver of G_1 contains at least $f(D, n)$ edges, and at least one of them contains precisely $f(D, n)$ edges. Consider the D-preserver G_1' of G_1 that contains precisely $f(D, n)$ edges. For any pair of vertices $u, w \in V_1$ such that $d_{G_1}(u, w) \geq D$, $d_{G_1'}(u, w) = d_{G_1}(u, w)$. Consider some subgraph $G_1'' = (V_1, E_1'')$ of G_1' such that E_1'' is a strict subset of E_1' (i.e., $E_1'' \subset E_1'$). As $|E_1''| < |E_1'| = f(D, n)$, it follows that G_1'' is not a D-preserver of G_1. I.e., there is a pair of vertices $u, w \in V_1$ such that $d_{G_1}(u, w) \leq D$, but $d_{G_1''}(u, w) > d_{G_1}(u, w) = d_{G_1'}(u, w)$. Hence, G_1'' is not a D-preserver of G_1' as well. Hence any D-preserver of G_1' contains at least $f(D, n)$ edges. As $\bar{f}(D, n)$ is the maximal number of edges in a graph whose any D-preserver contains at least the same number of edges as the graph itself, it follows that $f(D, n) \leq \bar{f}(D, n)$. This concludes the proof. ∎

Analogously, $f_S(D, n) = \bar{f}_S(D, n)$, $f^{dir}(D, n) = \bar{f}^{dir}(D, n)$ and $f_S^{dir}(D, n) = \bar{f}_S^{dir}(D, n)$. Also, as any subgraph D-preserver is, in particular, a Steiner D-preserver, it follows that $f_S(D, n) = \bar{f}_S(D, n) \leq f(D, n) = \bar{f}(D, n)$, and $f_S^{dir}(D, n) = f_S^{dir}(D, n) \leq f^{dir}(D, n) = \bar{f}^{dir}(D, n)$.

4.2 Lower Bounds

4.2.1 Undirected Graphs

The following example shows that for $0 < \delta < 1$ there is no fixed $D = D(\delta)$ such that for any undirected n-vertex graph G there exists a D-preserver G' with $O(n^{1+\delta})$ edges. Consider a clique of $n^{1/2+\delta/2}$ vertices (in this extended abstract we ignore the issue of a possible non-integrality of different quantities; anyway this affects only the lower order terms), with a path of length $D = n^{1/2-\delta/2}$ attached to every vertex. Denote this graph by $G_0 = (V_0, E_0)$.

Lemma 4.4 $f(D, n) = \bar{f}(D, n) = \Omega(n^2/D^2)$.

Proof: Let $W = \{w_1, w_2, \ldots, w_{n/D}\}$ be the set of the vertices of the clique, and $U = \{u_1, u_2, \ldots, u_{n/D}\}$ be the set of the endpoints of the paths that do not belong to the clique. Assume also that w_i's and u_i's are ordered in such a way that for any $i = 1, 2, \ldots, n/D$, w_i and u_i are two endpoints of the same path of length D.

Note that $|E_0| = \Theta(n^{1+\delta}) = \Theta(n^2/D^2)$. Also, observe that no strict subgraph of G_0 may serve as a D-preserver for G_0. This is because removing an edge from one of the paths makes the graph disconnected. In particular, in this case the distance between the non-clique endpoint of the path from which the edge was removed, and an endpoint of some other path, becomes infinity, and it is $2D - 1 \geq D$ in G_0. Also, removal of some sclique edge (w_i, w_j), $i \neq j$, $i, j = 1, 2, \ldots, n/D$ results in increasing the distance between u_i and u_j. Note that $d_{G_0}(u_i, u_j) \geq 2D$. Hence, $\bar{f}(D, n) = \Omega(n^2/D^2)$. Therefore, by Lemma 4.3, $f(D, n) = \Omega(n^2/D^2)$. ▮

Note that $f(D, n) = \Omega(n^2/D^2)$ and $f(D, n) = O(n^{1+\delta})$ implies $D = \Omega(n^{1/2-\delta/2})$. In other words, for any $0 < \delta < 1$, there are n-vertex graphs for which any subgraph with $O(n^{1+\delta})$ edges is not a D-preserver, for any $D = o(n^{1/2-\delta/2})$.

Note, however, that the graph G_0 does admit a *Steiner* 1-preserver of linear size. In this Steiner graph $V_0' = V_0 \cup \{s\}$, and the clique of size n/D in G_0 is replaced in G_0' by a star rooted in the new vertex s. All the edges of this star are of weight $1/2$. The paths remain unchanged.

Next, we show that

$$\bar{f}_S(D, n) \geq n^2/4D \ . \tag{8}$$

This improves the lower bound of Lemma 4.4 in two respects. First, this lower bound applies to *Steiner* D-preservers, while the lower bound of Lemma 4.4 applies only to subgraph D-preservers. Second, this lower bound is stronger by a factor of $\Theta(D)$ than that of Lemma 4.4.

Consider the following example. Let $G_1 = (V_1, E_1)$ be an $n/2 \times n/2D$ complete bipartite graph between the vertex sets $X = \{x_1, x_2, \ldots, x_{n/2}\}$ and $Y = \{y_1, y_2, \ldots, y_{n/2D}\}$ with paths of length $(D - 1)$ attached to each y_i, that connect y_i with z_i for $i = 1, 2, \ldots, n/2D$. It is easy to see that the only subgraph D-preserver of G_1 is G_1 itself. As the graph contains $|E| \geq n^2/4D$ edges, a lower bound of $f(D, n) = \bar{f}(D, n) \geq n^2/4D$ follows. Let $\vec{G}_1$ be the digraph obtained by replacing every edge of G_1 by two arcs, one in each direction. As the only subgraph D-preserver of $\vec{G}_1$ is $\vec{G}_1$, a lower bound on $f^{dir}(D, n)$ follows:

$$f^{dir}(D, n) = \bar{f}^{dir}(D, n) \geq n^2/2D \ . \tag{9}$$

However, the analogous lower bound for Steiner D-preservers applies only to the undirected case and requires a more delicate treatment (it is easy to see that $\vec{G}_1$ admits a *directed* Steiner 1-preserver with linear number of edges).

Consider an (undirected) Steiner D-preserver $G_1' = (V_1', E_1', \omega)$ of G_1. Assume, without loss of generality, that $\omega(e) > 0$ for an edge $e \in E'$. (Recall that by the definition of a Steiner graph, $\omega(e) \geq 0$.) Indeed, consider an edge $e = (u, w)$ such that $\omega(e) = 0$. First, note that either $u \in V_1' \backslash V_1$ or $w \in V_1' \backslash V_1$ (or both of them). This is because if $u, w \in V_1$ then $d_{G_1'}(u, w) \geq d_{G_1}(u, w)$, by definition of Steiner graph. Therefore, the edge (u, w) can be contracted (and if one of the vertices belongs to V_1, then the other one is eliminated) without changing the distances between the pairs of vertices $s, t \in V_1$.

In addition, for every pair $(i, j) \in \{1, 2, \ldots, n/2\} \times \{1, 2, \ldots, n/2D\}$, let us associate a shortest path $P_{i,j}$ between x_i and z_j.

Next, we describe Procedure *Extract* that will be used later on in the proof of inequality (8). The procedure accepts as input a graph G', and returns nothing. However, throughout the proof we will refer to the values of different variables in different stages of the execution of the procedure. The input graph of the procedure will be a Steiner D-preserver G'_1 of G_1. The procedure initializes the set UE of *unused edges* to contain the entire edgeset E' of G'. It also initializes the set UP of *uncovered pairs* to contain all possible pairs $\{(i,j) \mid i = 1,2,\ldots,n/2,\ j = 1,2,\ldots,n/2D\}$. The sets of *used edges*, CE, and covered pairs, CP, are both initialized to be empty sets. The main loop of the procedure runs while there is at least one uncovered pair (i,j). Inside the main loop, the procedure initializes the set of edges covered in this iteration, CE_0, to be an empty set, picks an uncovered pair (i,j), removes it from the set of uncovered pairs UP, inserts it into the set of covered pairs CP, removes the edgeset of $P_{i,j}$ from the set of unused edges UE and inserts it into the set CE_0. Then throughout the internal loop, the procedure looks for pairs (i',j') such that edgesets of $P_{i',j'}$ share at least one edge with CE_0, that is, with one of the other paths that were picked through the same iteration of the main loop. Upon finding such a pair, the procedure inserts it into CP, removes it from UP, inserts its edgeset into CE_0 and removes it from UE. The procedure leaves the internal loop whenever all the paths $P_{i',j'}$ that correspond to uncovered pairs (i',j') share no edge with CE_0. The main loop continues while not all the pairs are covered.

The idea of the proof is to associate with each pair (i,j) an edge $e' \in E'$ of the Steiner D-preserver G' via an injective mapping that is defined implicitly by Procedure *Extract*. It will follow that $|E'| \geq |\{1,2,\ldots,n/2\} \times \{1,2,\ldots,n/2D\}| = n^2/4D$. The formal description of Procedure *Extract* follows.

Procedure *Extract*

1. $UE \leftarrow E'$; $UP \leftarrow \{(i,j) \mid i = 1,2,\ldots,n/2,\ j = 1,2,\ldots,n/2D\}$; $CE, CP, CE_0 \leftarrow \emptyset$;
2. **While** $((UP \neq \emptyset)$ and $(UE \neq \emptyset))$ **do**
   begin (steps 3-5)
3. Pick $(i,j) \in UP$; set $UP \leftarrow UP \backslash \{(i,j)\}$; $CP \leftarrow CP \cup \{(i,j)\}$; $CE_0 \leftarrow CE_0 \cup E'(P_{i,j})$; $UE \leftarrow UE \setminus E'(P_{i,j})$;
4. **While** $\exists (i',j') \in UP$ s.t. $CE_0 \cap E'(P_{i',j'}) \neq \emptyset$ **do**
   $CP \leftarrow CP \cup \{(i',j')\}$; $UP \leftarrow UP \backslash \{(i',j')\}$; $CE_0 \leftarrow CE_0 \cup E'(P_{i',j'})$; $UE \leftarrow UE \backslash E'(P_{i',j'})$;
5. $CE \leftarrow CE \cup CE_0$; $CE_0 \leftarrow \emptyset$;
   end

Consider an execution of an invocation *Extract*(G') for some Steiner D-preserver G' of G_1. Let k be the number of iterations of the main loop during the invocation. Note that k is finite as in every iteration of the main loop at least one pair (i,j) is eliminated from UP. For $l = 1,2,\ldots,k$ let UP_l, CP_l, UE_l and CE_l be the values of the variables UP, CP, UE and CE at the beginning of the lth iteration. Also, let UP_{k+1}, CP_{k+1}, UE_{k+1} and CE_{k+1} be the values of these variables at the end of kth iteration. In addition, let $\widehat{CP}_l = CP_{l+1} \setminus CP_l$ and $\widehat{CE}_l = CE_{l+1} \setminus CE_l$.

Consider some fixed execution of an invocation *Extract*(G'). This execution can be divided into disjoint time periods, one time period for each step of the execution. Let $t_1, t_2, t_3, \ldots$ be points on the axis of time, where after t_p time units p steps of the execution were already completed, and $(p+1)$st step still did not start. Let $UP(p)$, $CP(p)$, $UE(p)$, $CE(p)$ and $CE_0(p)$ be the values of the variables UP, CP, UE, CE and CE_0 after t_p time units. For $l = 1,2,\ldots,k$, let j_l be the index such that the lth iteration of the main loop starts after t_{j_l} time units. In particular,

$UP_l = UP(t_{j_l})$, $CP_l = CP(t_{j_l})$, $UE_l = UE(t_{j_l})$ and $CE_l = CE(t_{j_l})$, for $l = 1, 2, \ldots, k$. The next lemmas illustrate some properties of these quantities.

Lemma 4.5 *For $p = 1, 2, \ldots$,*

$$UP(p) \cup CP(p) = \{(i,j) \mid i = 1,2,\ldots,n/2,\ j = 1,2,\ldots,n/2D\}\ , \tag{10}$$
$$UE(p) \cup CE(p) \cup CE_0(p) = E'\ . \tag{11}$$

Proof: By induction on p. The induction base ($p = 1$) follows from Step 1.

For the induction step, assume (10) and (11) for some $p = 1, 2, \ldots$ If during the interval $[t_p, t_{p+1}]$, the step that was executed affected no variable among UP, CP, UE, CE and CE_0 then the assertion follows from the induction hypothesis. Hence, it remains to consider the steps 3, 4 and 5. In the steps 3 and 4 whatever is inserted into CE_0 is removed from UE, and whatever is inserted into CP is removed from UP. In step 5 whatever is inserted into CE is removed from CE_0. Hence, the assertion follows from the induction hypothesis. ▮

Lemma 4.6 *For any $l = 1, 2, \ldots$, and any pair $(i,j) \in UP_l$, $E'(P_{i,j}) \subseteq UE_l$.*

Proof: Suppose for contradiction that there exists a pair $(i,j) \in UP_l$ such that $E'(P_{i,j}) \not\subseteq UE_l$. By Lemma 4.5, and as in the beginning of every iteration $CE_0 = \emptyset$, it follows that $E'(P_{i,j}) \subseteq UE_l \cup CE_l = E'$. Hence, $E'(P_{i,j}) \cap CE_l \neq \emptyset$. Let e be an edge in $E'(P_{i,j}) \cap CE_l$. Observe that as $CE_1 = \emptyset$,

$$\begin{aligned} CE_l &= CE_{l-1} \cup \widehat{CE}_{l-1} = CE_{l-2} \cup \widehat{CE}_{l-2} \cup \widehat{CE}_{l-1} \\ &= CE_1 \cup \bigcup_{k=1}^{l-1} \widehat{CE}_k = \bigcup_{k=1}^{l-1} \widehat{CE}_k\ . \end{aligned}$$

Hence, there exists an index $k = 1, 2, \ldots, l-1$ such that $e \in \widehat{CE}_k$. Hence, e was inserted into CE on the kth iteration of the main loop of the invocation *Extract*(G'). This could happen only on step 5. Hence e was inserted into CE_0 before the execution left the internal while loop (step 4) on the kth iteration of the main loop. Note also that $(i,j) \in UP(p)$, for any step $p \in [t_{j_k}, t_{j_{k+1}-1}]$. Hence, on the step when the execution left the internal while loop on the kth iteration of the main loop, the edge e was in CE, the pair (i,j) was in UP. Recall also that $e \in E'(P_{i,j})$. But this contradicts the exit condition of the internal while loop. Hence, for any pair $(i,j) \in UP_l$, $E'(P_{i,j}) \subseteq UE_l$. ▮

Corollary 4.7 *For any two distinct indices $l_1, l_2 = 1, 2, \ldots, k$, $l_1 \neq l_2$, $\widehat{CP}_{l_1} \cap \widehat{CP}_{l_2} = \emptyset$. $\widehat{CE}_{l_1} \cap \widehat{CE}_{l_2} = \emptyset$.*

Proof: The first assertion of the corollary follows directly from the fact that all pairs that are inserted into CP are drawn out of UP (i.e., belong to UP at the time of insertion into CP, and are removed from UP at the same time as they are inserted into CP). For the second assertion, note that by Lemma 4.6, all edges that are inserted into CE are drawn out of UE (i.e., belong to UE at the time of insertion into CE). ▮

The next next lemma shows that $\widehat{CE}_l$ has a very convenient structure.

Lemma 4.8 *Let $G' = (V', E', \omega)$ be a Steiner D-preserver of G and let $l = 1, 2, \ldots, k$. Then $\widehat{CE}_l = \bigcup_{P \in \Pi} E'(P)$, where $\Pi = \{P_{i_1,j}, P_{i_2,j}, \ldots, P_{i_r,j}\}$, $\{i_1, i_2, \ldots, i_r\}$ is an r-subset of $\{1, 2, \ldots, n/2\}$, and $j \in \{1, 2, \ldots, n/2D\}$.*

Proof: First, note that $\widehat{CE}_l = \bigcup_{P \in \Pi} E'(P)$ for $\Pi \subseteq \{P_{i,j} \mid (i,j) \in \{1, 2, \ldots, n/2\} \times \{1, 2, \ldots, n/2D\}\}$. It remains to prove that for every pair of paths $P_{i,j}, P_{i',j'} \in \Pi$, $j = j'$. Consider a subset $\widehat{CE} \subseteq \widehat{CE}_l$ that was formed on the $l$the iteration of the main loop after $p = 0, 1, 2, \ldots$ executions of the internal loop were completed (this is the value of the variable CE_0 after p iterations of the internal loop on the lth iteration of the main loop). Observe that $\widehat{CE} = \bigcup_{P \in \hat{\Pi}} E'(P)$ for some subset $\hat{\Pi} \subseteq \Pi$.

Let us show by induction on p that for any $P_{i,j}, P_{i',j'} \in \hat{\Pi}$, the indices j and j' are equal. To start the induction, note that whenever $p = 0$, the set $\hat{\Pi}$ contains a single path. For the induction step, assume the induction hypothesis for some p. Let $P_{i',j'}$ be the path whose edgeset is added into $\widehat{CE}_l$ in the $(p+1)$st iteration of the internal loop. Let $\widehat{CE}'$ be the value of the variable CE after the pth iteration of the internal loop. By the exit condition of the internal loop, $\widehat{CE} \cap E'(P_{i',j'}) \neq \emptyset$. Hence there exists a path $P_{i,j} \in \hat{\Pi}'$, where $\widehat{CE}' = \bigcup_{P \in \hat{\Pi}'} E'(P)$, such that $E'(P_{i,j}) \cap E'(P_{i',j'}) \neq \emptyset$. Note that $\widehat{CE} = \bigcup_{P \in \hat{\Pi}} E'(P)$, $\hat{\Pi} = \hat{\Pi}' \cup \{(i', j')\}$, and by the induction hypothesis, for every path $P_{i'',j''} \in \hat{\Pi}'$, $j'' = j$. Let $e \in E'(P_{i,j}) \cap E'(P_{i',j'})$. Let w be the closer endpoint of $e = (u, w)$ to z_j. Then, as G' is a D-preserver of G_1, and $d_{G_1}(x_i, z_j) = D$, and $\omega(e) > 0$,

$$d_{G'}(z_j, w) \;\leq\; d_{G'}(x_i, z_j) - \omega(e) \;=\; d_{G_1}(x_i, z_j) - \omega(e) \;=\; D - \omega(e) < D\ .$$

Furthermore,

$$d_{G'}(z_j, z_{j'}) \;\leq\; d_{G'}(z_j, w) + d_{G'}(w, z_{j'}) \;<\; D + d_{G'}(w, z_{j'}) \;\leq\; 2D\ .$$

(The last inequality is because w lies on $P_{i',j'}$, which is the shortest path in G' between $x_{i'}$ and $z_{j'}$; note also that $d_{G'}(x_{i'}, z_{j'}) = d_G(x_{i'}, z_{j'}) = D$.) Hence, $d_{G'}(z_j, z_{j'}) < 2D$. It follows that $j = j'$, as otherwise $d_{G'}(z_j, z_{j'}) < d_G(z_j, z_{j'}) = 2D$. ∎

This structure of $\widehat{CE}_l$ enables to derive the following inequality.

Lemma 4.9 *Consider a Steiner D-preserver $G' = (V', E', \omega)$ of G_1. Let $l = 1, 2, \ldots, k$, and $\widehat{CE}_l = \bigcup_{P \in \Pi} E'(P)$. Then $|\widehat{CE}_l| \geq |\Pi|$.*

Proof: By induction on $|\Pi|$. The induction base is $|\Pi| = 1$. Let $\Pi = \{P_{i,j}\}$. Then $\widehat{CE}_l = E'(P_{i,j})$. As $P_{i,j}$ is a path between two different vertices x_i and z_j, it follows that $|E'(P_{i,j})| \geq 1$, completing the proof of the induction base.

For the induction step, recall that by Lemma 4.8, $\Pi = \{P_{i_1,j}, P_{i_2,j}, \ldots, P_{i_r,j}\}$. By the induction hypothesis, $|\bigcup_{p=1}^{r-1} E'(P_{i_p,j})| \geq r - 1$. Recall that $P_{i_r,j}$ is a path between x_{i_r} and z_j. Note that $x_{i_r} \notin \{x_{i_1}, x_{i_2}, \ldots, x_{i_{r-1}}\}$. Also, x_{i_r} is not an internal vertex of $P_{i_p,j}$ for some $p = 1, 2, \ldots, r-1$. Indeed, otherwise $d_{G'}(x_{i_r}, z_j) < d_{G'}(x_{i_p}, z_j)$, but G' is a D-preserver of G, and so $d_G(x_{i_r}, z_j) = d_G(x_{i_p}, z_j) = D = d_{G'}(x_{i_r}, z_j) = d_{G'}(x_{i_p}, z_j)$. This is a contradiction. Hence $x_{i_r} \notin \bigcup_{p=1}^{r-1} V'(P_{i_p,j})$. But $x_{i_r}, z_j \in V'(P_{i_r,j})$.

Let u be the closest vertex to x_{i_r} that belongs to the set $\left(\bigcup_{j=1}^{r-1} V'(P_{i_p,j})\right) \cap V'(P_{i_r,j})$. Note that $x_{i_r} \neq u$, because $x_{i_r} \notin \bigcup_{j=1}^{r-1} V'(P_{i_p,j})$. Let P denote the subpath of $P_{i_r,j}$ between x_{i_r} and u. Observe that

$$E'(P) \cap \left(\bigcup_{j=1}^{r-1} E'(P_{i_p,j})\right) = \emptyset \ .$$

Also,$|E'(P)| \geq 1$. Hence

$$|\bigcup_{p=1}^{r} E'(P_{i_p,r})| \ \geq \ |\bigcup_{p=1}^{r-1} E'(P_{i_p,r}) \cup E'(P)| \ = \ |\bigcup_{p=1}^{r-1} E'(P_{i_p,r})| + |E'(P)| \ \geq \ |\bigcup_{p=1}^{r-1} E'(P_{i_p,r})| + 1 \ .$$

By induction hypothesis, $|\bigcup_{p=1}^{r-1} E'(P_{i_p,r})| \geq r-1$. Hence, $|\bigcup_{p=1}^{r} E'(P_{i_p,r})| \geq (r-1)+1 = r$. ▮

Corollary 4.10 *For any $l = 1, 2, \ldots, k$, $|\widehat{CE}_l| \geq |\widehat{CP}_l|$.*

Proof: Note that whenever $\widehat{CE}_l = \bigcup_{P \in \Pi} E'(P)$, $\widehat{CP}_l = \{(i,j) \mid P_{i,j} \in \Pi\}$. Hence, $|\widehat{CP}_l| = |\Pi|$. Now the assertion follows from Lemma 4.9. ▮

Suppose for contradiction that $|E'| < n^2/4D$. The following lemma holds under this assumption.

Lemma 4.11 *For a Steiner D-preserver $G' = (V', E', \omega)$ with $|E'| < n^2/4D$, and $l = 1, 2, \ldots, k$, $|UP_l| > |UE_l|$.*

Proof: By induction on l. For the induction base note that

$$|UP_1| \ = \ |\{(i,j) \mid i = 1, 2, \ldots, n/2, j = 1, 2, \ldots, n/2D\}| \ = \ n^2/4D > |E'| = |UE_1| \ .$$

For the induction step, assume for some $l = 1, 2, \ldots, k-1$ that $|UP_l| > |UE_l|$. Note that $UP_l = UP_1 \setminus \bigcup_{j=1}^{l-1} \widehat{CP}_j$, and $\widehat{CP}_l \cap (\bigcup_{j=1}^{l-1} \widehat{CP}_j) = \emptyset$. As $\widehat{CP}_l \subseteq UP_1$, it follows that $\widehat{CP}_l \subseteq UP_l$, and so $|UP_{l+1}| \ = \ |UP_l \setminus \widehat{CP}_l| \ = \ |UP_l| - |\widehat{CP}_l|$. Analogous consideration using Corollary 4.10 for $\widehat{CE}$ implies that $|UE_{l+1}| = |UE_l \setminus \widehat{CE}_l| = |UE_l| - |\widehat{CE}_l|$.

By the induction hypothesis, $|UP_l| > |UE_l|$, and by Corollary 4.10, $|\widehat{CP}_l| \leq |\widehat{CE}_l|$. Hence, $|UP_{l+1}| = |UP_l| - |\widehat{CP}_l| > |UE_l| - |\widehat{CE}_l| = |UE_{l+1}|$. ▮

To summarize,

Theorem 4.12 *For any $n = 2, 3, \ldots$ and $D = 1, 2, \ldots, n-1$, $f_S(D,n) = \bar{f}_S(D,n) \geq n^2/4D$.*

Proof: Recall that UP_{k+1} and UE_{k+1} are the values of the variables UP and UE, respectively, at the time of leaving the main loop of the invocation *Extract*(G'). Note that either $UP_{k+1} = \emptyset$ or $UE_{k+1} = \emptyset$. As by Lemma 4.11, $|UP_{k+1}| > |UE_{k+1}|$, it follows that $UE_{k+1} = \emptyset$ and $UP_{k+1} \neq \emptyset$. Recall that $\emptyset \ = \ UE_{k+1} \ = \ UE_k \setminus \widehat{CE}_k \ = \ UE_{k-1} \setminus (\widehat{CE}_k \cup \widehat{CE}_{k-1}) \ = \ \ldots \ = \ UE_1 \setminus \bigcup_{l=1}^{k} \widehat{CE}_l$. Hence $E' = UE_1 = \bigcup_{l=1}^{k} \widehat{CE}_l$. Let $(i,j) \in UP_{k+1}$. Note that $E'(P_{i,j}) \subseteq E' = \bigcup_{l=1}^{k} \widehat{CE}_l$. Hence there exists an index $l = 1, 2, \ldots, k$ such that $E'(P_{i,j}) \cap \widehat{CE}_l \neq \emptyset$. However, this contradicts the assumption that the invocation *Extract*(G') left the internal loop (step 4) on the lth iteration of the main loop. This is a contradiction to the assumption that $|UE_1| = |E'| < n^2/4D = |UP_1|$. Therefore, in any Steiner D-preserver $G' = (V', E', \omega)$ of the n-vertex graph $G_1 = (V_1, E_1)$, $|E'| \geq n^2/4D$. ▮

4.2.2 Directed Graphs and Distance-Preserving Labeling Schemes

We next turn to proving a lower bound on $\bar{f}_S^{dir}(D,n) = f_S^{dir}(D,n)$.

Consider again the digraph $\vec{G}_1$ mentioned in Section 4.2.1. Recall that the only subgraph D-preserver of $\vec{G}_1 = (V_1, \vec{E}_1)$ is the digraph itself (see inequality (9) and, also, any *undirected* Steiner D-preserver of this graph requires $\Omega(n^2/D)$ edges. However, as we mentioned, this digraph does admit a *directed* Steiner 1-preserver $G'_1 = (V'_1, E'_1, \omega)$ with *linear* number of edges. Specifically, $V'_1 = V_1 \cup \{s_l, s_r\}$. Every vertex $x \in X \subseteq V_1$ is connected via an outgoing arc $\langle x, s_l \rangle$ to s_l, and via an incoming arc $\langle s_r, x \rangle$ to s_r. Also, every vertex $y \in Y \subseteq V_1$ is connected via an incoming arc $\langle s_l, y \rangle$ to to s_l, and via an outgoing arc $\langle y, s_r \rangle$ to s_r. All these arcs are of weight 1/2. The paths between y_i and z_i for $i = 1, 2, \ldots, n/2D$ are not modified. It is easy to see that for every pair of vertices $u, w \in V_1$, $d_{G'_1}(u,w) = d_{\vec{G}_1}(u,w)$. Also, $|E'_1| \leq 3/2n + n/D$. Hence, the digraph $\vec{G}_1$ cannot serve as an example that shows that $f_S^{dir}(D,n) = \Omega(n^2/D)$. Furthermore, we will show in Section 4.3 that this claim is not true, and $f_S^{dir}(D,n) = O(\frac{n^2 \log D}{D \log n})$. In particular, it will follow that for $D = O(1)$, for any digraph there is a *directed Steiner*, referred later on as a *diSteiner*, D-preserver with $O(n^2/\log n)$ arcs, where all the arcs are of weight 1 or 1/2. This separates the directed case from the undirected one, as $f_S(D,n) = \Omega(n^2/D)$ (see Theorem 4.12). Generalizing this upper bound, it will be shown there that for any digraph with $O(n^2/2^{\log^\gamma n})$ arcs, $0 < \gamma < 1$, a factor of $\Theta(\frac{\log^{1-\gamma} n}{\log\log n})$ can be "saved" using a diSteiner 1-preserver. Furthermore, some constant factor can be "saved" all the way to $n^{3/2}$. We next argue that there are n-vertex graphs G with $m = \Omega(n^{3/2})$ arcs such that any diSteiner 1-preserver of G contains at least m arcs.

Let $G = (U, W, E)$ be a bipartite graph with girth greater than 4. In other words, G contains no subgraph isomorphic to $K_{2,2}$.

We next argue that every diSteiner 1-preserver of G contains at least $|E|$ arcs.

Lemma 4.13 *Let $G' = (V', E', \omega)$ be a diSteiner 1-preserver of G. Then $|E'| \geq |E|$.*

Proof: Let G' be a diSteiner 1-preserver of the bipartite graph $G = (U, W, E)$. It follows that for any edge $e = (u, w) \in E$ there exists a path $P_e = P_{u,w}$ in G' of length 1. Associate such a path P_e with every edge $e \in E$ (if there are several such paths, pick one of them arbitrarily). We next argue that

$$|\bigcup_{e \in E} E'(P_e)| \geq |E| \ .$$

This would imply $|E'| \geq |E|$, as $|E'| \geq |\bigcup_{e \in E} E'(P_e)|$.

Consider an arbitrary ordering $(e_1, e_2, \ldots, e_{|E|})$ of the edges of E. Let $\mathcal{E}_k = \bigcup_{i=1}^{k} E'(P_{e_i})$.

Lemma 4.14 $|\mathcal{E}_k| \geq k$*, for* $k = 1, 2, \ldots, |E|$.

Proof: The proof is by induction on k. For the induction base ($k = 1$), note that $|\mathcal{E}_1| = |E'(P_{e_1})| \geq 1$.

Assume the induction hypothesis for some $k = 1, 2, \ldots, |E| - 1$. It remains to argue that $|\mathcal{E}_{k+1} \setminus \mathcal{E}_k| \geq 1$. Let $e_{k+1} = (u, w)$. Let $\mathcal{E}(u,w) = \{(u', w') \in \mathcal{E}_k \mid E'(P_{u',w'}) \cap E'(P_{u,w}) \neq \emptyset\}$. Observe that for any edge $(u', w') \in \mathcal{E}(u,w)$, either $u = u'$ or $w = w'$. Indeed, otherwise let $s \in V'(P_{u,w}) \cap V'(P_{u',w'})$. Denote by $P_{u,s}$ (resp., $P_{u',s}$) the subsegment of $P_{u,w}$ (resp., $P_{u',w'}$) from u (resp., u') to s, and by $P_{s,w}$ (resp., $P_{s,w'}$) the subsegment of $P_{u,w}$ (resp., $P_{u',w'}$) from s to w (resp.,

w'). Note that $1 = |P_{u,w}| = |P_{u,s}| + |P_{s,w}| = |P_{u',w'}| = |P_{u',s}| + |P_{s,w'}|$. Suppose for contradiction that $|P_{u,s}| < |P_{u',s}|$. But then $d'_G(u,w') \leq |P_{u,s}| + |P_{s,w'}| < |P_{u',s}| + |P_{s,w'}| = |P_{u',w'}| = 1 \leq d_G(u,w')$, i.e., $d_{G'}(u,w') < d_G(u,w')$, contradiction. The assumption $|P_{u',s}| < |P_{u,s}|$ yields a constradiction in an analogous way. Hence, $|P_{u',s}| = |P_{u,s}|$.

It follows that $d_{G'}(u,w) = d_{G'}(u',w') = d_{G'}(u,w') = d_{G'}(u',w) = 1 = d_G(u,w) = d_G(u',w') = d_G(u,w') = d_G(u',w)$. I.e., $(u,w),(u',w'),(u,w'),(u',w) \in E$, contradicting the assumption that no $K_{2,2}$ is contained in G.

So, for any edge $(u',w') \in \mathcal{E}(u,w)$ either $u = u'$ or $w = w'$. Note also that as $(u,w) \notin \mathcal{E}_k$, it follows that $(u,w) \notin \mathcal{E}(u,w)$, and thus either $u \neq u'$ or $w \neq w'$. Let $\mathcal{E}^u(u,w) = \{(u',w') \in \mathcal{E}(u,w) \mid u = u'\}$ and $\mathcal{E}^w(u,w) = \{(u',w') \in \mathcal{E}(u,w) \mid w = w'\}$. As we argued $\mathcal{E}(u,w) = \mathcal{E}^u(u,w) \cup \mathcal{E}^w(u,w)$, and, $\mathcal{E}^u(u,w) \cap \mathcal{E}^w(u,w) = \emptyset$.

We next define a total order relation $\leq_v$ of the vertices of $V'(P_{u,w})$ as follows. For a pair of vertices $x, y \in V'(P_{u,w})$, $x \leq_v y$ if and only if $d_{G'}(u,x) \leq d_{G'}(u,y)$.

Observe that for any edge $(u,w') \in \mathcal{E}^u(u,w)$, its corresponding path $P_{u,w'}$ "branches out" of the path $P_{u,w}$ at some point. Let $s(w')$ be the biggest vertex in $V'(P_{u,w}) \cap V'(P_{u,w'})$ with respect to the order relation $\leq_v$. We also define a total order relation $\leq_e$ on the edges of $\mathcal{E}^u(u,w)$ as follows. For a pair of edges $(u,w_1),(u,w_2) \in \mathcal{E}^u(u,w)$, $(u,w_1) \leq_e (u,w_2)$ if and only if $s(w_1) \leq_v s(w_2)$.

Analogously, for any edge $(u',w') \in \mathcal{E}^w(u,w)$, let $s(u')$ be the smallest vertex of $V'(P_{u,w}) \cap V'(P_{u',w})$ with respect to the order relation $\leq_e$. The total order relation $\leq_e$ on the edges of $\mathcal{E}^w(u,w)$ is defined in an analogous way.

Let (u,w') be the biggest edge in $\mathcal{E}^u(u,w)$, and (u',w) be the smallest edge in $\mathcal{E}^w(u,w)$ (both with respect to the order relation $\leq_e$; if there are several biggest edges, pick arbitrarily one of them).

Observe that by definition of $\mathcal{E}^u(u,w)$ and $\mathcal{E}^w(u,w)$, u, u', w, w' are distinct vertices of $V(G)$. Let $s(w')$ be the biggest vertex of $V'(P_{u,w}) \cap V'(P_{u,w'})$, and $s(u')$ be the smallest vertex of $V'(P_{u,w}) \cap V'(P_{u',w})$. It follows that $s(u') >_v s(w')$, as otherwise it would follow that the vertices u, u', w and w' form $K_{2,2}$ in G, and this is a contradiction. Let $P_{s(w'),s(u')}$ denote the subsegment of $P_{u,w}$ between $s(w')$ and $s(u')$. It remains to argue that

$$E'(P_{s(w'),s(u')}) \cap \bigcup_{e \in \mathcal{E}_k} E'(P_e) = \emptyset \ . \tag{12}$$

Indeed, suppose for contradiction that there exists an edge $e \in \mathcal{E}_k$ such that $E'(P_e) \cap E'(P_{s(w'),s(u')}) \neq \emptyset$. It follows that $e \in \mathcal{E}(u,w) = \mathcal{E}^u(u,w) \cup \mathcal{E}^w(u,w)$. Recall that $\mathcal{E}^u(u,w) \cap \mathcal{E}^w(u,w) = \emptyset$. Hence $e \in \mathcal{E}^u(u,w)$ or $e \in \mathcal{E}^w(u,w)$.

Consider the case $e \in \mathcal{E}^u(u,w)$ (the case is $e \in \mathcal{E}^w(u,w)$ is analogous). Then $e = (u,w'')$ for some $w'' \in W$. Observe that as $E'(P_e) \cap E'(P_{s(w'),s(u')}) \neq \emptyset$, $s(u'), s(w') \in V'(P_e) \cap V'(P_{s(w'),s(u')})$, and so there exists a vertex $z \neq s(w')$ such that $z \in V'(P_e) \cap V'(P_{s(w'),s(u')})$. Note that $z \in V'(P_e)$, and $s(w') <_v z$. Observe also that $z \leq_v s(w'')$. It follows that $s(w') <_v s(w'')$, and so $(u,w') <_e (u,w'')$, contradicting the assumption that the edge (u,w') is the biggest in $\mathcal{E}^u(u,w)$ with respect to the total order $\leq_e$. Now (12) follows. ∎

This completes the proof of Lemma 4.13. ∎

Corollary 4.15 *There are n-vertex digraphs G with $m \geq (1/2 + o(1))n^{3/2}$ edges such that any diSteiner 1-preserver of G contains at least m arcs.*

Proof: As demonstrated in [6], there are bipartite graphs G_0 with $(1/2 + o(1))n^{3/2}$ edges with $girth(G_0) > 4$. The corollary follows by orienting all its arcs consistently from one bipartition to another, and using Lemma 4.13. ∎

In what follows we show that $\bar{f}_S^{dir}(D,n) = f_S^{dir}(D,n) = \Omega(\frac{n^2 \log D}{D \log n})$.

Let $\mathcal{G}$ be the family of graphs with a common vertex set V. The vertex set V is comprised of $X = \{x_1, x_2, \ldots, x_{n/2}\}$, $Y = \{y_1, y_2, \ldots, y_{n/(4D)}\}$, $Z = \{z_1, z_2, \ldots, z_{n/4D}\}$ and vertices of the paths connecting y_j to z_j for every $j = 1, 2, \ldots, n/4D$, $2D-2$ vertices apart of y_j and z_j in each path. For every graph $G \in \mathcal{G}$, its edgeset contains the paths of length $2D-1$ from y_j to z_j for every $j = 1, 2, \ldots, n/4D$. For every $j = 1, 2, \ldots, n/4D$ and $l = 1, 2, \ldots, 2D-1$, let y_j^0 denote y_j, and y_j^l denote the vertex that is on distance l from y_j, and is located on the path connecting y_j and z_j. (In particular, $y_j^{2D-1} = z_j$.) In addition, for every $i = 1, 2, \ldots, n/2$, $j = 1, 2, \ldots, n/4D$, G contains precisely one arc from x_i to y_j^l, for some $l = 0, 1, \ldots, D-1$. All the arcs are unit-weight. The family $\mathcal{G}$ consists of all the digraphs G that can be constructed this way.

It follows that

$$|\mathcal{G}| = D^{n/2 \cdot n/4D} = 2^{\frac{n^2 \log D}{8D}} . \tag{13}$$

We need the following definition.

Definition 4.16 *The graph G' is a (D,g)-preserver of $G = (V,E)$ if for every pair of vertices $u, w \in V$ such that $d_G(u,w) \geq D$, $d_G(u,w) \leq d_{G'}(u,w) \leq d_G(u,w) + g$.*

Lemma 4.17 *Let G_1' and G_2' be Steiner $(D, 1/3n)$-preservers of two distinct n-vertex graphs $G_1, G_2 \in \mathcal{G}$. Then $G_1' \neq G_2'$.*

Proof: As $G_1 \neq G_2$, there exists a pair $(i,j) \in \{1, 2, \ldots, n/2\} \times \{1, 2, \ldots, n/4D\}$ such that $\langle x_i, y_j^{l_1} \rangle \in E(G_1)$, $\langle x_i, y_j^{l_2} \rangle \in E(G_2)$, and $l_1 \neq l_2$. For these i and j, $|d_{G_1}(x_i, z_j) - d_{G_2}(x_i, z_j)| \geq 1$. Observe also that as $l_1, l_2 \leq D-1$, it follows that $d_{G_1}(x_i, z_j), d_{G_2}(x_i, z_j) \geq (2D-1) - (D-1) + 1 = D+1$. It follows that $|d_{G_1'}(x_i, z_j) - d_{G_2'}(x_i, z_j)| \geq |d_{G_1}(x_i, z_j) - d_{G_2}(x_i, z_j)| - 2/3n = 1 - 2/3n > 0$, for any $n = 1, 2, \ldots$. Hence, $d_{G_1'}(x_i, z_j) \neq d_{G_2'}(x_i, z_j)$. It follows that $G_1' \neq G_2'$. ∎

Fix n, and consider the family $\mathcal{G}$ of n-vertex digraphs discussed above. Let $V = (v_1, v_2, \ldots, v_n)$ be the vertex be an arbitrary ordering of the (common to all graphs of $\mathcal{G}$) vertex set V. For a distance labeling scheme $(\mathcal{M}, \mathcal{D})$, and a graph $G \in \mathcal{G}$, let $\mathcal{M}(G) = \mathcal{M}(G, v_1) \cdot \mathcal{M}(G, v_2) \cdot \ldots \cdot \mathcal{M}(G, v_n)$, where "$\cdot$" stands for concatenation.

Lemma 4.18 *Let $(\mathcal{M}, \mathcal{D})$ be a distance-labelling D-preserving scheme and $G_1, G_2 \in \mathcal{G}$, $G_1 \neq G_2$. Then $\mathcal{M}(G_1) \neq \mathcal{M}(G_2)$.*

Proof: Similarly to the proof of Lemma 4.17, since $G_1 \neq G_2$, there exists a pair of vertices $x_i, z_j \in V$ such that $d_{G_1}(x_i, z_j), d_{G_2}(x_i, z_j) \geq D$, and $d_{G_1}(x_i, z_j) \neq d_{G_2}(x_i, z_j)$.

As $(\mathcal{M}, \mathcal{D})$ is a D-preserving scheme, it follows that $\mathcal{D}(\mathcal{M}(G_1, x_i), \mathcal{M}(G_1, z_j)) = d_{G_1}(x_i, z_j)$ and $\mathcal{D}(\mathcal{M}(G_2, x_i), \mathcal{M}(G_2, z_j)) = d_{G_2}(x_i, z_j)$. Hence,
$\mathcal{D}(\mathcal{M}(G_1, x_i), \mathcal{M}(G_1, z_j)) \neq \mathcal{D}(\mathcal{M}(G_2, x_i), \mathcal{M}(G_2, z_j))$. Hence, either $\mathcal{M}(G_1, x_i) \neq \mathcal{M}(G_2, x_i)$ or $\mathcal{M}(G_1, z_j) \neq \mathcal{M}(G_2, z_j)$ (or both). In either case, $\mathcal{M}(G_1) \neq \mathcal{M}(G_2)$. ∎

Let φ be an arbitrary representation function of the Steiner $(D, 1/3n)$-preservers of graphs from the family $\mathcal{G}$. Specifically, with each graph $G \in \mathcal{G}$, φ associates a bit string of length k, that determines uniquely some specific Steiner $(D, 1/3n)$-preserver G' of G. Note that by Lemma 4.17, φ is injective. Indeed, if $G' = \varphi(G_1) = \varphi(G_2)$ then G' is a Steiner $(D, 1/3n)$-preserver of both G_1 and G_2, and so, by Lemma 4.17, $G_1 = G_2$. Hence, by (13),

Corollary 4.19 *For every representation function of the Steiner $(D, 1/3n)$-preservers of $\mathcal{G}$ there exists a graph $G \in \mathcal{G}$ such that $|\varphi(G)| \geq \log |\{\varphi(G) \mid G \in \mathcal{G}\}| = \log |\mathcal{G}| = \frac{n^2 \log D}{8D}$.*

Analogously, Lemma 4.18 implies a lower bound on D-preserving distance labeling schemes. Note that all the lower bounds in this section apply both to the directed and undirected graphs. However, for undirected Steiner graphs stronger lower bounds were shown in Section 4.2.1. This is not the case for the distance labeling schemes, where the lower bound below is the strongest that we are able to prove.

Corollary 4.20 *Every distance labeling D-preserving scheme requires labels of size $\Omega(\frac{n \log D}{D})$ bits.*

Intuitively, the last stage of the proof of the lower bound $f_S^{dir}(D, n) = \Omega(\frac{n^2 \log D}{D \log n})$ is proving that using non-rational (or even rational but having very large denominator) weights cannot help saving arcs of the diSteiner D-preservers. This is done in the next theorem. The technique of getting rid of the non-rational weights in a Steiner graph, that is used in the proof, is adapted from [1], where Steiner spanners with a multiplicative approximation of distances are studied.

Theorem 4.21 *For $n = 2, 3, \ldots,$ the family of n-vertex digraphs $\mathcal{G}$ defined above, and $D = 1, 2, \ldots, n-1$, let $\rho : \mathcal{G} \to \mathcal{G}'$ be a function assigning to every digraph $G \in \mathcal{G}$ a diSteiner D-preserver G'. Then there exists a digraph $G \in \mathcal{G}$ such that $G' = \rho(G)$ contains $\Omega(\frac{n^2 \log D}{D \log n})$ arcs.*

Proof: Consider a mapping $\rho' : \mathcal{G}' \to \mathcal{G}''$ that given a digraph $G' = (V', E', \omega)$ constructs a digraph $G'' = (V', E', \omega')$, where for every arc $e \in E'$, $\omega(e)$ is defined to be the closest rational number with denominator $1/3n^3$. Let $\rho'' : \mathcal{G} \to \mathcal{G}''$ be the composition of ρ and ρ'.

Suppose for contradiction that for any digraph $G \in \mathcal{G}$, its diSteiner D-preserver $G' = \rho(G)$ contains less than $\frac{n^2 \log D}{6 \cdot (8D \log n)}$ arcs. In particular, it follows that for any digraph $G \in \mathcal{G}$, its diSteiner D-preserver $G' = (V', E', \omega)$ has at most n^2 vertices. Hence for any pair of vertices $u, w \in V'$, any simple path from u to w in G' contains no more than n^2 arcs. As for every arc $e \in E'$, $|\omega(e) - \omega'(e)| \leq 1/3n^3$, it follows that for any simple path P from u to w in G', $|\omega(P) - \omega(P')| \leq n^2/3n^3 = 1/3n$.

As G' is a diSteiner D-preserver of G, it follows that $\rho'(G') = G''$ is a diSteiner $(D, 1/3n)$-preserver of G. Observe also that for any $G \in \mathcal{G}$, the digraphs $G' = \rho(G)$ and $G'' = \rho''(G)$ have the same arcset. By our assumption, for every digraph $G \in \mathcal{G}$, $G' = \rho(G)$ contains less than $\frac{n^2 \log D}{6 \cdot (8D \log n)}$ arcs. It follows that for every digraph $G \in \mathcal{G}$, $G'' = \rho''(G)$ contains less than $\frac{n^2 \log D}{6 \cdot (8D \log n)}$ arcs. Observe also that for any arc $e \in E(G'')$, its weight in G'' is rational number. As all the distances in G are no greater than $n - 1$, and G'' is a diSteiner $(D, 1/3n)$-preserver, we assume, without loss of generality, that all the arcs in G'' have weight that is no greater than n. Hence, every arc $e \in E(G'')$ can be represented by a bit string $\alpha(e)$ of length $6 \log n$, by writing down the identities of its endpoints ($2 \log n$ bits), and the numerator of its weight (at most $\log n^4 = 4 \log n$ bits).

The representation function φ is now formed out of ρ by concatenating in an arbitrary but fixed order the strings $\alpha(e)$ for different arcs $e \in E(G'')$. Observe that for any digraph $G \in \mathcal{G}$, $\varphi(G)$ determines uniquely a diSteiner $(D, 1/3n)$-preserver G'' of G, and $|varphi(G)|$ contains $\frac{n^2 \log D}{6\cdot(8D\log n)} 6\log n = \frac{n^2 \log D}{8D \log n}$ bits. However, this contradicts Corollary 4.19.

Hence there is a digraph $G \in \mathcal{G}$ such that its diSteiner D-preserver $\rho(G) = G'$ contains at least $\frac{n^2 \log D}{48D \log n}$ arcs. ∎

4.2.3 (D, g)-Preservers

To facilitate the discussion about (D, g)-preservers, we generalize Definition 4.2 in the following way.

Definition 4.22 *For $n = 2, 3, \ldots,$ and $D, g = 1, 2, \ldots, n-1$, let $f(D, g, n)$ be the minimal number such that for any n-vertex graph there exists a (D, g)-preserver with at most $f(D, g, n)$ edges, and let $\bar{f}(D, g, n)$ be the maximal number of edges in an n-vertex graph whose only subgraph (D, g)-preserver is the graph itself.*

The following "weak duality" follows directly from the definition.

Lemma 4.23 *For $n = 2, 3, \ldots,$ and $D, g = 1, 2, \ldots, n-1$, we have $f(D, g, n) \geq \bar{f}(D, g, n)$.*

However, unlike the case with no additive error, no upper bound on $f(D, g, n)$ in terms of $\bar{f}(D, g, n)$ is known to the authors.

We next show a lower bound on $\bar{f}(D, g, n)$, which serves, consequently, as a lower bound on $f(D, g, n)$.

Theorem 4.24 *For $D, g = 1, 2, \ldots,$ and n sufficiently large, $f(D, g, n) \geq \bar{f}(D, g, n) \geq \frac{n^{1+c_0/(g+2)}}{2g \cdot D^{c_0/(g+2)}}$, where c_0 is some constant $1 \leq c_0 \leq 2$.*

Remark: *(The lower bound on the size of an extremal n-vertex graph of girth g stands currently on $\Omega(n^{1+c_0/(g-1)})$ [5], for $c_0 = 4/3$. Erdős conjectured that $c_0 = 2$.)*

Proof: Set $L = \lfloor n/2D \rfloor$. $L = n/2D$. There exists a constant $1 \leq c_0 \leq 2$ such that there exists an L-vertex graph $G_0 = (V_0, E_0)$ with $girth(G_0) \geq g + 2$ and $|E_0| \geq L^{1+c_0/(g+2)}$ (cf. [19], p.166). Denote the vertices of G_0 by the numbers $1, 2, \ldots, L$. (I.e., $V_0 = \{1, 2, \ldots, L\}$.)

To build the graph $G^{(D,g)}$, we begin with L paths of length D: vertices v_{ij}, $i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, D$, and edges $(v_{ij}, v_{i,j+1})$, $i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, D-1$.

Add $L \cdot D/(g/2)$ vertices w_{ij}, $i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, D/(g/2)$, and for any $i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, D/(g/2)$ connect v_{i1} to w_{ij} by a path of length $g/2$.

For each j, $j = 1, 2, \ldots, D/(g/2)$, construct an isomorphic copy of G_0 using the vertices $\{w_{ij}\}_{i=1}^{L}$. Specifically, for each j, $j = 1, 2, \ldots, D/(g/2)$, for every $i, h = 1, 2, \ldots, L$, add the edge (w_{ij}, w_{hj}) if and only if $(i, j) \in E_0$.

The number of vertices is $L \cdot (D + g/2 \cdot D/(g/2)) = 2LD \leq 2D\lfloor n/(2D) \rfloor \leq n$; add $n - 2DL$ vertices to one of the paths to absorb the slack, giving $G^{(D,g)}$ exactly n vertices (i.e., $|V^{(D,g)}| = n$). $L \cdot (D + g/2 \cdot D/(g/2)) = 2LD = n$.

The number of the edges is

$$
\begin{aligned}
|E^{(D,g)}| &\geq L\cdot(D-1) + (L\cdot D/(g/2))\cdot g/2 + n - 2DL + \lfloor n/2D \rfloor^{1+c_0/(g+2)}\cdot D/(g/2) \\
&\geq n - \lfloor n/2D \rfloor + \frac{n^{1+c_0/(g+2)}\cdot 2D}{2^{1+c_0/(g+2)}D^{c_0/(g+2)}g} \geq \frac{n^{1+c_0/(g+2)}}{2gD^{c_0/(g+2)}} .
\end{aligned}
$$

Let us argue that $G^{(D,g)}$ is the only (D,g)-preserver of itself.

Indeed, removing a path edge $e = (v_{ij}, v_{i+1,j})$ for some $i = 1,2,\ldots,D-1$, $j = 1,2,\ldots,L$ makes the graph disconnected, and, in particular, $d_G(w_{ij'}, v_{iD}) \geq D$, and $d_{G_e}(w_{ij'}, v_{iD}) = \infty$, for any $j' = 1,2,\ldots,D/(g/2)$.

Removing an edge from a path that connects v_{i1} with w_{ij} for some $i = 1,2,\ldots,L$, $j = 1,2,\ldots,D/(g/2)$ increases the distance between w_{ij} and v_{iD} by at least $g+1$.

Finally, removing an edge (w_{ij}, w_{hj}) increases the distance from w_{hj} to v_{iD} by at least $g+1$, since for any $j = 1,2,\ldots,D/(g/2)$, the graph $G^{(D,g)}(\{w_{ij} \mid i = 1,2,\ldots,L\})$ has girth equal to $g+2$. ▮

4.3 Upper bounds

4.3.1 Distance Preservers

We start with presenting an almost matching (up to a constant factor of 4) upper bound on the size of possible distance D-preservers.

Lemma 4.25 *For $n = 2,3,\ldots$ and $D = 1,2,\ldots,n-1$,*

$$f^{dir}(D,n) = \bar{f}^{dir}(D,n) \leq 2n(n-1)/(D+1) , \quad f(D,n) = \bar{f}(D,n) \leq n(n-1)/(D+1) .$$

Proof: Suppose that for any arc $e \in E$, the path $P(e) = P(e,D)$ exists.

Consider some vertex $v \in V$. We next argue that for any two arcs that are outgoing from v, $e_1 = \langle v, z_1\rangle$, $e_2 = \langle v, z_2\rangle$,

$$V(\mathit{suffix}(P(e_1), e_1)) \cap V(\mathit{suffix}(P(e_2), e_2)) = \emptyset .$$

$V(\mathit{suffix}(P(e_1), e_1)) \cap V(\mathit{suffix}(P(e_2), e_2)) = \emptyset$.
Suppose for contradiction that some vertex $w \in V(\mathit{suffix}(P(e_1), e_1)) \cap V(\mathit{suffix}(P(e_2), e_2))$. By (2),
Then, $d_{G_{e_1}}(\mathit{head}(P(e_1)), \mathit{tail}(P(e_1))) > d_G(\mathit{head}(P(e_1)), \mathit{tail}(P(e_1)))$, and
$d_{G_{e_2}}(\mathit{head}(P(e_2)), \mathit{tail}(P(e_2))) > d_G(\mathit{head}(P(e_2)), \mathit{tail}(P(e_2)))$.

For $i = 1,2$, let P'_i, P''_i and P'''_i be the segments of $P(e_i)$ from $\mathit{head}(P(e_i))$ to v, from v to w, and from w to $\mathit{tail}(P(e_i))$, respectively. See Figure 1.

Note that since $P(e_1)$ is the shortest path between $\mathit{head}(P(e_1))$ and $\mathit{tail}(P(e_1))$ in G,
$d_G(\mathit{head}(P(e_1)), \mathit{tail}(P(e_1))) = |P'_1| + |P''_1| + |P'''_1|$.

Consider the walk $P_{12} = P'_1 \cdot P''_2 \cdot P'''_1$. Note that P_{12} is a walk between $\mathit{head}(P_1)$ and $\mathit{tail}(P_1)$ in $E \setminus \{e_1\}$. Hence,

$$
\begin{aligned}
|P'_1| + |P''_2| + |P'''_1| &\geq d_{G_{e_1}}(\mathit{head}(P(e_1)), \mathit{tail}(P(e_1))) \\
&> d_G(\mathit{head}(P(e_1)), \mathit{tail}(P(e_1))) = |P'_1| + |P''_1| + |P'''_1| .
\end{aligned}
$$

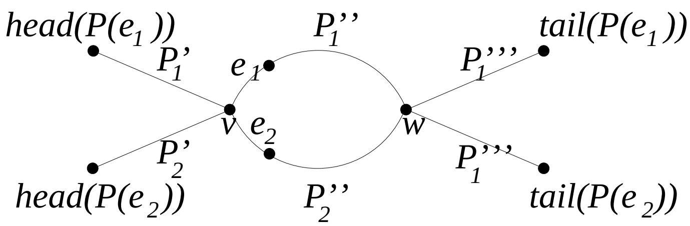

Figure 1: The subpaths of $P(e_1)$ and $P(e_2)$.

Hence $|P_2''| > |P_1''|$. However, analogously, it follows that $|P_1''| > |P_2''|$, contradiction.

Therefore, the set $\mathcal{P}_{out}(v) = \{\mathit{suffix}(P(\langle v, z\rangle), \langle v, z\rangle) \mid \langle v, z\rangle \in E\}$ consists of vertex-disjoint paths.

Analogously, it follows that the set $\mathcal{P}_{in}(v) = \{\mathit{prefix}(P(\langle z, v\rangle), \langle z, v\rangle) \mid \langle z, v\rangle \in E\}$ consists of vertex-disjoint paths.

Note that for every vertex $v \in V$ and path $P \in \mathcal{P}_{in}(v) \cup \mathcal{P}_{out}(v)$, the node v does not belong to $V(P)$. Thus,

$$\sum_{P \in \mathcal{P}_{out}(v)} |V(P)|, \sum_{P \in \mathcal{P}_{in}(v)} |V(P)| \le |V \setminus \{v\}| = n - 1 \ .$$

Thus,

$$\sum_{v \in V} \left(\sum_{P \in \mathcal{P}_{in}(v)} |V(P)| + \sum_{P \in \mathcal{P}_{out}(v)} |V(P)| \right) \le 2n(n-1) \ .$$

$\sum_{v \in V} \left(\sum_{P \in \mathcal{P}_{in}(v)} |V(P)| + \sum_{P \in \mathcal{P}_{out}(v)} |V(P)| \right) \le 2n(n-1)$.
Also, since for any arc $e \in E$, $|P(e)| \ge D$,

$$\begin{aligned}
& \sum_{v \in V} \left(\sum_{P \in \mathcal{P}_{in}(v)} |V(P)| + \sum_{P \in \mathcal{P}_{out}(v)} |V(P)| \right) \\
= & \sum_{v \in V} \left(\sum_{\langle z,v\rangle \in E} |V(\mathit{prefix}(P(\langle z, v\rangle), \langle z, v\rangle))| + \sum_{\langle v,z\rangle \in E} |V(\mathit{suffix}(P(\langle v, z\rangle), \langle v, z\rangle))| \right) \\
= & \sum_{\langle v,z\rangle \in E} (|V(\mathit{prefix}(P(\langle v, z\rangle), \langle v, z\rangle))| + |V(\mathit{suffix}(P(\langle v, z\rangle), \langle v, z\rangle))|) \\
= & \sum_{\langle v,z\rangle \in E} (|\mathit{prefix}(P(\langle v, z\rangle), \langle v, z\rangle)| + \mathit{1} + |\mathit{suffix}(P(\langle v, z\rangle), \langle v, z\rangle)| + \mathit{1}) \\
= & \sum_{\langle v,z\rangle \in E} (|P(\langle v, z\rangle)| + 1) \ = \sum_{\langle v,z\rangle \in E} |P(\langle v, z\rangle)| + |E| \ \ge \ |E| \cdot D + |E| \ .
\end{aligned}$$

Thus, $|E| \cdot (D + 1) \le 2n(n-1)$.

For an undirected graph $G = (V, E)$, the analogous argument provides an upper bound which is smaller by a factor of 2. ∎

Note that the inequalities in Lemma 4.25 are tight for $D = 1$, since there is a graph (n-vertex clique K_n) with $n \cdot (n-1)/(D+1) = n \cdot (n-1)/2$ edges, in which removal of any edge results in increasing the distance between some pair of vertices that are already at distance at least $D = 1$. Also, there is a digraph (complete n-vertex digraph) with $2n \cdot (n-1)/(D+1) = n \cdot (n-1)$ arcs, with the same property.

The next theorem indicates that the product $D \cdot f(D, n)$ is independent of D and equal to $\Theta(n^2)$.

Theorem 4.26 (Distance×Size Preservation)
For $n = 2, 3, \ldots$ and $D = 1, 2, \ldots, n-1$,

$$n^2/4D \;\leq\; f_S(D,n) \;\leq\; f(D,n) \;\leq\; n(n-1)/(D+1)\ , \tag{14}$$
$$n^2/2D \;\leq\; f^{dir}(D,n) \;\leq\; 2n(n-1)/(D+1)\ . \tag{15}$$

Proof: Both upper bounds follow from Lemma 4.25. The lower bound of inequality (14) follows from Theorem 4.12. The lower bound of inequality (15) follows from (9). $f^{dir}(D,n) = \bar{f}^{dir}(D,n) \geq n^2/2D$. ▮

We next prove a tight up to a constant factor upper bound on $f_S^{dir}(D, n)$.

Consider an n-vertex digraph $G = (V, E)$ with $m = \Omega(n^{3/2})$ arcs. Suppose $V = \{1, 2, \ldots, n\}$. The digraph G can be represented by its $n \times n$ adjacency matrix $M(G)$, whose entry (i, j) is 1 if and only if $\langle i, j \rangle \in E$, and 0 otherwise. Suppose, without loss of generality, that the digraph contains no loops (that is, arcs $\langle i, i \rangle$ for some $i \in V$) as the latter can be removed from the digraph with no affect on the distances. Set $c'' = 1 + \nu_1$ for some arbitrarily small positive constant $\nu_1 > 0$. Denote $p = m/(c''n^2)$.

Lemma 4.27 *$M(G)$ contains an $a \times a$ submatrix containing all 1's with $a = \lfloor c' \log n / \log(1/p) \rfloor$, for $c' = 1 - \nu_2$ for some arbitrarily small positive constant $\nu_2 > 0$.*

Remark: Such a matrix corresponds to $K_{a,a}$, that is, complete bipartite subgraph of size $a \times a$ with all arcs oriented consistently from one bipartition of the subgraph to another.
Proof: Following Zarankiewicz, let us denote by $k_a(n)$ the least number m such that any n-vertex digraph G with at least m arcs contains a $K_{a,a}$. The assertion of the lemma is a corollary of the following result from [18], chapter 5.

Theorem 4.28 *[18] If $n\binom{m/n}{a} \geq (a-1)\binom{n}{a}$ then $k_a(n) \leq m$.*

To show that the assumption of Theorem 4.28 is satisfied, it is enough to argue that

$$n \cdot \left(\frac{m/n}{n} \cdot \frac{m/n - 1}{n-1} \cdot \ldots \cdot \frac{m/n - a + 1}{n - a + 1} \right) \;\geq\; a\ .$$

As $m/n = \Omega(\sqrt{n})$ and $a = O(\log n)$, it follows that for any sufficieanly large n and any $i = 1, 2, \ldots, a-1$,

$$\frac{m/n - i}{n - i} \geq \frac{m/(c''n)}{n}\ .$$

Hence, it is sufficient to argue that $n(m/(c''n))^a = n \cdot p^a \geq a$. Substituting $a = c' \log n / \log(1/p)$ implies $n^{1-c'} \geq a$, and the latter is true for sufficiently large n (as $a = O(\log n)$). Theorem 4.28 implies $k_a(n) \leq m$. The assertion of the lemma now follows from the definition of $k_a(n)$. ∎

Let $m_0 = m = n^2/D$ for some D be the number of arcs in $G_0 = G$, and $p_0 = p = m_0/(c''n^2)$ be the "density" of the arcs. Set $\epsilon = \frac{\log D}{\log n}$ (i.e., $D = 2^{\epsilon \log n}$). Set also $S_0 = 0$ to be the number of arcs inserted into the diSteiner graph so far. By Lemma 4.27, G contains a subgraph isomorphic to K_{a_0,a_0} with $a_0 = c' \log n / (\log 1/p_0)$. Pick such a subgraph and represent it with a diSteiner vertex s (in addition to $2a_0$ original vertices) and $2a_0$ appropriately oriented arcs of weight 1/2 each connecting s with the original vertices. The orientation of these arcs is the following: all the arcs between s and "left-hand" vertices (those that had only outgoing arcs in the chosen subgraph) are incoming into s, and all the other arcs are out-going from s. The constructed structure is inserted into the diSteiner graph, and the charge S is updated from $S_0 = 0$ to $S_1 = S_0 + 2a_0 = 2a_0$. Delete the arcs of chosen subgraph from G_0, and denote the obtained digraph G_1. The density p changes according to $p_1 = p_0 - a_0^2/(c''n^2)$. If the number of arcs in G_1 is still greater than $\mu \cdot \frac{n^2(\log D + \log e)}{D \cdot \log n}$ for some arbitrarily small constant $\mu > 0$, repeat this procedure with $a_1 = c' \log n / (\log 1/p_1)$. Observe that the condition on the number of arcs implies that $a_1 \geq 1$, and so in a finite number r of iterations we are left with a digraph G_r with at most $\mu \cdot \frac{n^2(\log D + \log e)}{D \cdot \log n}$ arcs. When the number of arcs left is at most $\mu \cdot \frac{n^2(\log D + \log e)}{D \cdot \log n}$, these arcs are inserted into the diSteiner graph G'.

Lemma 4.29 *The constructed digraph G' is a diSteiner 1-preserver of G.*

Proof: Consider some arc $\langle u, w \rangle \in E$. Either at one of the iterations e was replaced by two arcs $\langle u, s \rangle, \langle s, w \rangle$ of weight 1/2 each, for some new vertex s, or the arc e was inserted into G'. In either case $d_{G'}(u, w) = d_G(u, w) = 1$. It follows that for any pair of vertices $x, y \in V(G)$, $d_{G'}(x, y) \leq d_G(x, y)$.

Also, it can be shown by induction on r that for any $x, y \in V(G)$, $d_G(x, y) \leq d_{G'}(x, y)$. Intuitively, this is because whenever an isomorphic to a $K_{a,a}$ between $x_1, x_2, \ldots, x_a$ and $y_1, y_2, \ldots, y_a$ is replaced by a star of arcs $\langle x_1, s \rangle, \langle x_2, s \rangle, \ldots, \langle x_a, s \rangle, \langle s, y_1 \rangle, \langle s, y_2 \rangle, \ldots, \langle s, y_a \rangle$, no paths between x_i and x_j or y_i and y_j are formed. This is unlike the undirected case, where such a replacement could cause $d_{G'}(x_i, x_j) < d_G(x_i, x_j)$. This is, however, quite natural, as in the undirected case there are graphs for which any Steiner 1-preserver contains $\Omega(n^2)$ edges (see Theorem 4.26, inequality (14)).

It follows that G' is a diSteiner 1-preserver of G. ∎

Next, we calculate the number of arcs in G'.

Lemma 4.30 *If n is sufficiently large then*

$$S_r \leq \frac{2c''}{c' - \epsilon} \cdot \frac{n^2}{D} \cdot \frac{\log D + \log e}{\log n} . \quad (16)$$

Proof: Observe that $S_r = S_0 + 2\sum_{i=0}^{r-1} a_i = 2\sum_{i=0}^{r-1} a_i$. Denote $\Delta p_i = p_{i+1} - p_i$ for $i = 0, 1, \ldots, r-1$. Note that $\Delta p_i > 0$ for $i = 0, 1, \ldots, r-1$. Then $S_r/2 = \sum_{i=0}^{r-1} \frac{a_i}{\Delta p_i} \Delta p_i$. Observe that $\Delta p_i = p_{i+1} - p_i = \frac{a_i^2}{c''n^2}$. Hence $\frac{a_i}{\Delta p_i} = c''n^2/a_i$. By Lemma 4.27, $a_i \geq c'(\frac{\log n}{\log 1/p_i} - 1/c')$. Substituting $p_i \geq \mu \frac{\log D}{c'' D \log n}$

and $D = 2^{\epsilon \log n}$ implies that $\log 1/p_i \le \epsilon \log n - \log \mu\epsilon$. Hence $\frac{\log n}{\log 1/p_i} - 1/c' \ge (1-\epsilon/c')\frac{\log n}{\log 1/p_i}$. Therefore, $a_i/\Delta p_i \le \frac{c''}{c'}\frac{1}{1-\epsilon/c'}\frac{n^2 \log 1/p_i}{\log n} = \frac{c''}{c'-\epsilon} \cdot \frac{n^2 \log 1/p_i}{\log n}$. Hence

$$S_r/2 \le c'' \cdot \frac{1}{c'-\epsilon} n^2/\log n \sum_{i=0}^{r-1} \log 1/p_i \Delta p_i \ . \tag{17}$$

Observe that as $p_0 > p_1 > \ldots p_{r-1} > p_r > 0$, it follows that $\sum_{i=0}^{r-1} \log 1/p_i \Delta p_i$ is a Riemann sum of $\int_0^{p_0} (\log 1/p) dp$. Furthermore, $\Delta p_i = a_i^2/(2n^2) \le \frac{\log^2 n}{n^2}$. Hence Δp_i tends to 0 when n grows, for any $i = 0, 1, 2, \ldots, r-1$. Hence for any $\delta > 0$ there exists a sufficiently large n such that

$$\sum_{i=0}^{r-1} \log 1/p_i \Delta p_i \ \le \ \int_0^{p_0} (\log 1/p) dp + \delta \ \le \ p_0(\log 1/p_0 + 1) + \delta \ .$$

Now, the lemma follows from (17). ∎

Corollary 4.31 *For every n-vertex (di)graph with m edges (resp., arcs) the following statements hold.*

1. *There exists a diSteiner 1-preserver with $O(n^2/\log n)$ arcs.*
2. *If $m \le n^2/\log^c n$ for some $c > 0$ then there exists a diSteiner 1-preserver with $O(\frac{c \cdot n^2 \log\log n}{\log^c n})$ arcs.*
3. *If $m \le n^{1+\alpha}$, $0 < \alpha < 1$, then there exists a diSteiner 1-preserver with at most $\frac{2+\mu}{\alpha}(1-\alpha)\cdot m$ arcs for any arbitrarily small constant μ.*
4. *There exists a diSteiner D-preserver with $O(\frac{n^2 \log D}{D \log n})$ arcs. I.e.,*

$$\bar{f}_S^{dir}(D,n) = f_S^{dir}(D,n) = \Theta(\frac{n^2 \log D}{D \log n}) \ .$$

The weights of arcs in the aforementioned diSteiner graphs may be restricted to be either 1 or 1/2.

Proof: For assertion (1), substitute $\epsilon = 0$ to Lemma 4.30. It follows that $S_r \le ((2+\nu)n^2/\log n$, for some arbitrarily small constant $\nu > 0$. The assertion follows as the number of arcs in the diSteiner 1-preserver is

$$S_r + \mu \cdot \frac{n^2 \cdot \log D}{D \cdot \log n} = (2+\mu)n^2/\log n$$

for an arbitrarily small constant $\mu > 0$.

The assertion (2) follows analogously, by substituting $\epsilon = c \log\log n/\log n$.

For assertion (3), note that $D = n^2/n^{1+\alpha} = n^{1-\alpha}$. I.e., $\epsilon = 1-\alpha$. Now the assertion follows from Lemma 4.30.

For assertion (4), recall that by Theorem 4.26, for any n-vertex (di)graph there exists a subgraph D-preserver with $O(n^2/D)$ edges (resp., arcs). If $D = \Omega(n^\epsilon)$ for some constant $\epsilon > 0$ then $O(n^2/D) = O(\frac{n^2 \log D}{D \log n})$. Otherwise, if $D = 2^{\epsilon(n)\cdot \log n}$ for some $\epsilon(n)$ such that $\lim_{n\to\infty} \epsilon(n) = 0$, then the assertion follows from Lemma 4.30, and from the observation that a 1-preserver of a D-preserver of a graph G is a D-preserver of G.

Finally, the lower bound

$$\bar{f}_S^{dir}(D,n) = f_S^{dir}(D,n) = \Omega\left(\frac{n^2 \log D}{D \log n}\right)$$

follows from Theorem 4.21. ▮

Note that by Corollary 4.31, for any graph with at least $m = n^{5/3+\delta}$ edges (for any $\delta > 0$) there exists a diSteiner 1-preserver with strictly less than m arcs. This statement can be generalized to $m \geq c \cdot n^{3/2}$, for some small constant $c > 1$, by extracting subgraphs isomorphic to $K_{s,2}$ for different decreasing values of s whenever no $K_{3,3}$ can be extracted. Note that the latter cannot be generalized much further, as by Corollary 4.15 there exist n-vertex graphs with $m = (1/2 + o(1))n^{3/2}$ edges for which any diSteiner 1-preserver contains at least m arcs.

4.3.2 Algorithmic Aspects

In this section we address some algorithmic aspects of our results concerning distance D-preservers. In particular, we devise a distance labeling D-preserving scheme with labels of size $O((n^2/D) \cdot \log^2 n)$. Recall that by Corollary 4.20 labels of size $O((n^2/D) \cdot \log D)$ are required.

Theorem 4.32 *For $n = 2, 3, \ldots$, $D = 1, 2, \ldots, n-1$, and an n-vertex graph (resp., digraph) with m edges (resp., arcs), there exists a constructible in $O(m^3 n)$ time subgraph D-preserver with at most $n(n-1)/(D+1)$ edges (resp., $2n(n-1)/(D+1)$ arcs).*

Proof: We prove the assertion for a digraph G; the proof of the slightly stronger statement for the undirected graphs is analogous.

The proof is by induction on the number of arcs in G, $|E| = m$. The induction base is $|E| \leq \frac{2n \cdot (n-1)}{D+1}$. In this case $G' = (V, H)$ with $H = E$ is the subgraph with the desired properties.

For the induction step, suppose that for any digraph G with $|E| = m \geq \frac{2n \cdot (n-1)}{D+1}$ arcs exists a subgraph $G' = (V, H)$, $H \subseteq E$ with the desired properties.

Consider a graph $\bar{G} = (\bar{V}, \bar{E})$ with $|\bar{E}| = m + 1$ arcs. Since $m + 1 > \frac{2n \cdot (n-1)}{D+1} \geq \bar{f}^{dir}(D, n)$, there exists an arc $e \in \bar{E}$ such that for any pair of vertices $u, w \in \bar{V}$ with $d_{\bar{G}}(u, w) \geq D$,

$$d_{\bar{G}_e}(u, w) = d_{\bar{G}}(u, w) \ . \qquad (18)$$

Note that the cardinality of the set of arcs of $\bar{G}_e$ is $|\bar{E} \setminus \{e\}| = |\bar{E}| - 1 = m$, and so the induction hypothesis is applicable to $\bar{G}_e$. In other words, there exists a subgraph $G' = (\bar{V}, H)$ of $\bar{G}_e$, $H \subseteq \bar{E} \setminus \{e\} \subseteq \bar{E}$, with $|H| \leq 2n(n-1)/(D+1)$, such that for any pair of vertices $u, w \in \bar{V}$ such that $d_{\bar{G}_e}(u, w) \geq D$,

$$d_{G'}(u, w) = d_{\bar{G}_e}(u, w) \ . \qquad (19)$$

Note that, by (18), $d_{\bar{G}_e}(u, w) \geq D$ implies $d_{\bar{G}}(u, w) = d_{\bar{G}_e}(u, w) \geq D$, and so, it follows that $G' = (\bar{V}, H)$ is a subgraph of $\bar{G} = (\bar{V}, \bar{E})$, $H \subseteq \bar{E}$, with $|H| \leq 2n(n-1)/(D+1)$, such that for any pair of vertices $u, w \in V$ with $d_{\bar{G}}(u, w) = d_{\bar{G}_e}(u, w) \geq D$, $d_{G'}(u, w) = d_{\bar{G}_e}(u, w) = d_{\bar{G}}(u, w)$. The last two equalities are by (18) and (19).

Note that the edge e as above can be found in polynomial time, by computing all the distances in $\bar{G}_e$ for every $e \in \bar{E}$, and testing whether there is a pair of vertices $u, w \in \bar{V}$ such that $d_{\bar{G}}(u,w) \geq D$ and $d_{\bar{G}_e}(u,w) > d_{\bar{G}}(u,w)$.

Therefore, the entire computation of the subgraph G', that satisfies the assertion of the theorem, can be completed in polynomial time (specifically, in $O(|E|^3 \cdot n)$ time). ▮

We remark that after inequalities (14) and (15) were communicated to Mikkel Thorup, he devised [23] a more efficient randomized procedure for computing a subgraph D-preserver of size $O(n^2 \log n/D)$ (greater than optimal by a logarithmic factor). This more efficient procedure uses some techniques of [26] from the area of dynamic algorithms. The efficiency of the procedure of [23] makes it more suitable for algorithmic applications such as (and this is, indeed, the motivation of [23]) computing shortest paths between pairs of vertices that are at distance at least D one from another. We next use a similar idea to prove the existence of a distance labeling D-preserving scheme with labels of size $O((n/D) \cdot \log^2 n)$. This is tight up to a factor of $O(\log^2 n / \log D)$, in view of Corollary 4.20.

Theorem 4.33 *For $D = 1, 2, \ldots$ there exists a distance labeling D-preserving scheme $(\mathcal{M}, \mathcal{D})$ for a family of all (possibly directed) n-vertex unweighted graphs with labels of size $O((n/D) \cdot \log^2 n)$.*

Proof: Fix $2 < c < 3$ be some real constant. Consider a labeling procedure that given an n-vertex graph $G = (V, E)$ starts with choosing a random subset $R \subseteq V$ of vertices. Every $v \in V$ is chosen into R independently at random with probability $p = \min\{c \log n/D, 1\}$.

Next, the procedure fixes an arbitrary ordering $(u_1, u_2, \ldots, u_{|R|})$ of the vertices of R. Then, for every pair of vertices $v \in V$, $u \in R$, the procedure forms a string $\alpha_v(u)$ to be the concatenation of the bit strings $d_G(v,u)$ and $d_G(u,v)$ (if the graph G is undirected, $\alpha(u)$ is the bit string representing $d_G(v,u) = d_G(u,v)$).

Finally, for every vertex $v \in V$, the procedure forms its label $\mathcal{M}(G,v)$ to be $\alpha_v(u_1) \cdot \alpha_v(u_2) \cdot \ldots \cdot \alpha_v(u_{|R|})$, where "$\cdot$" stands for concatenation.

Observe that $|\mathbb{E}(R)| = p \cdot n \leq c \log n \cdot n/D$. Hence, for every vertex $v \in V$, $|\mathcal{M}(G,v)| \leq c \log n \cdot n^2/D$. The query-answering procedure accepts as input two labels $\mathcal{M}(G,v_1) = \alpha_{v_1}(u_1) \cdot \alpha_{v_1}(u_2) \cdot \ldots \cdot \alpha_{v_1}(u_{|R|})$ and $\mathcal{M}(G,v_2) = \alpha_{v_2}(u_1) \cdot \alpha_{v_2}(u_2) \cdot \ldots \cdot \alpha_{v_2}(u_{|R|})$, and returns $\min\{d_G(v_1,u) + d_G(u,v_2) \mid u \in R\}$. Observe that for every $u \in R$, $d_G(v_i,u)$ can be computed given $\mathcal{M}(G,v_i)$, $i = 1, 2$.

By Markov inequality,

$$\mathbb{P}(|R| \leq 2c \log n \cdot n/D) \geq 1/2 \ . \tag{20}$$

For every pair of vertices (v_1, v_2), fix some shortest path P_{v_1,v_2} from v_1 to v_2 in G. (In an undirected graph P_{v_1,v_2} coincides with P_{v_2,v_1}.) Observe that for v_1, v_2 such that $d_G(v_1,v_2) \geq D$, $|V(P_{v_1,v_2})| \geq D + 1$. Note that for a vertex $z \in V(P_{v_1,v_2})$, $\mathbb{P}(z \in R) = c \log n/D$. Hence

$$\mathbb{P}(V(P_{v_1,v_2}) \cap R = \emptyset) \ = \ (1 - c \log n/D)^{D+1} \ \leq \ 1/n^c \ .$$

Hence,

$$\mathbb{P}(\exists v_1, v_2 \in V \ \text{ s.t. } \ d_G(v_1,v_2) \geq D \text{ and } V(P_{v_1,v_2}) \cap R = \emptyset) \ \leq \ n^2/n^c = 1/n^{c-2} \ .$$

I.e.,

$$\mathbb{P}(\forall v_1, v_2 \in V \ \text{ s.t. } \ d_G(v_1,v_2) \geq D, V(P_{v_1,v_2}) \cap R \neq \emptyset) \ \geq \ 1 - 1/n^{c-2} \ .$$

Together with (20), this implies that

$$\mathbb{P}(|R| \le 2c\log n \cdot n/D \text{ and } \forall v_1, v_2 \in V \text{ s.t. } d_G(v_1,v_2) \ge D, V(P_{v_1,v_2}) \cap R \ne \emptyset) \ \ge\ 1/2 - 1/n^{c-2} .$$

Finally, note that the event $(\forall v \in V, |\mathcal{M}(G,v)| \le 2c\log^2 n \cdot n/D)$ contains the event $(|R| \le 2c\log n \cdot n/D)$, and for every pair of vertices $v_1, v_2 \in V$ the event $(V(P_{v_1,v_2}) \cap R \ne \emptyset)$ contains the event $(\mathcal{D}(\mathcal{M}(G,v_1), \mathcal{M}(G,v_2)) = d_G(v_1,v_2)$. Hence,

$$\begin{aligned}&\mathbb{P}(\forall v \in V, |\mathcal{M}(G,v)| \le 2c\log^2 n \cdot n/D, \text{ and } \forall v_1, v_2 \in V \text{ s.t. } d_G(v_1,v_2) \ge D,\\ &\mathcal{D}(\mathcal{M}(G,v_1), \mathcal{M}(G,v_2)) = d_G(v_1,v_2)) \ \ge 1/2 - 1/n^{c-2} \ >\ 0 ,\end{aligned}$$

for sufficiently large n.

Hence, there exists a D-preserving distance labeling scheme with labels of size $O(\log^2 n \cdot n/D)$. ∎

Next, we devise a polynomial time algorithm for constructing a diSteiner 1-preserver with $O(n^2/\log n)$ arcs for an arbitrary graph. In conjunction with Theorem 4.32, this yields a polynomial time algorithm for constructing a diSteiner D-preserver with $O(\frac{n^2\log D}{D\log n})$ arcs for an arbitrary graph.

We remark that the main obstacle towards converting the proof of Corollary 4.31 into an efficient algorithm is the existential nature of the proof of Theorem 4.28. Next theorem is a constructive proof version of Theorem 4.28, Lemma 4.27. that is, an efficient algorithm for extracting a subgraph isomorphic to $K_{s,t}$ from a sufficiently dense graph. Another algorithm with a similar running time for extracting $K_{s,t}$ was devised by [16], and our algorithm is provided for completeness.

For any vertex $y \in V$, let $d(y)$ denote the degree of y.

Theorem 4.34 *[16] Let G be a graph of order n, $W \subseteq V(G)$, and $1 \le s,t$. Suppose*

$$\sum_{y\in W} \binom{d(y)}{t} > (s-1)\binom{n}{t} . \tag{21}$$

Then G contains a $K_{s,t}$ with the 's part' contained in W, i.e., there are (necessarily disjoint) sets $S \subset W$ and $T \subset V$, $|S| = s$, $|T| = t$, such that every vertex of S is joined to every vertex of T. The $K_{s,t}$ can be computed in $O(n^2 \cdot t)$ time.

Proof: We shall do considerably more than claimed by the theorem: we shall give an algorithm that finds a 'large' set $S \subset W$ all whose vertices are joined to all vertices of a set T with t vertices. Our condition (21) will imply that the set S constructed by the algorithm will have at least s vertices.

In our description of the algorithm, we shall say that a triple (G,W,t), with $W \subset V(G)$, is *s-large*, if condition (21) is satisfied.

Here is then our plan. Starting with the triple (G,W,t), we perform the *t-step* of the algorithm to construct a vertex x_1 and a triple $(G_1, W_1, t-1)$, where $G_1 = G - x_1$, $W_1 \subset W \setminus \{x_1\}$, the vertex x_1 is joined to all vertices in W_1, and the triple $(G_1, W_1, t-1)$ is s-large, then perform the $(t-1)$-step of the algorithm to obtain a vertex $x_2 \in V(G_1)$ and a triple $(G_2, W_2, t-2)$ with $G_2 = G_1 - x_2$ and $W_2 \subset W_1 \setminus \{x_2\}$, such that x_2 is joined to every vertex in W_2, and the triple $(G_2, W_2, t-2)$ is s-large, and so on. Finally, after the 1-step of the algorithm, we get a vertex x_t and a triple $(G_t, W_t, 1)$. This completes the algorithm: our sets are $S = W_t$ and $T = \{x_1, x_2, \ldots, x_t\}$. By construction, G contains all edges from S to T and, as $(G_t, W_t, 1)$ is s-large, from (21) $\sum_{y\in W}\binom{d(y)}{t} > (s-1)\binom{n}{t}$ we shall find that $|S| \geq s$.

To complete our proof, here is then the t-step of the algorithm. For $x \in V(G)$, let the (t, W)-weight of x be

$$w(x) = w_{t,W}(x) = \sum_{(x,y)\in E,\ y\in W} \binom{d(y)-1}{t-1} .$$

Since

$$\sum_{x\in V} w(x) = \sum_{y\in W} d(y)\binom{d(y)-1}{t-1} = \sum_{y\in W} t\binom{d(y)}{t} > t(s-1)\binom{n}{t} = (s-1)n\binom{n-1}{t-1} ,$$

there is a vertex $x_1 \in V$ such that

$$\sum_{y\in W_1} \binom{d(y)-1}{t-1} > (s-1)\binom{n-1}{t-1}, \tag{22}$$

where $W_1 = \{y \in W : (x, y) \in E(G)\}$. Indeed, any vertex whose (W, t)-weight is at least the average will do for x_1; a vertex of maximal (W, t)-weight will certainly do. Set $G_1 = G - x_1$. Condition (22) means precisely that the triple $(G_1, W_1, t-1)$ is s-large (as for any $y \in W_1$, its degree in G_1 is $d(y) - 1$). Hence we can apply the $(t-1)$-step of our algorithm to the triple $(G_1, W_1, t-1)$, and so on, until we get to an s-large triple $(G_t, W_t, 1)$. Since

$$|W_t| = \sum_{y\in W_t} \binom{d(y)-1}{0} > (s-1)\binom{n-1}{0} = s-1,$$

we find that $|W_t| \geq s$. By construction, the graph G contains all edges from $S = W_t$ to $T = \{x_1, x_2, \ldots, x_t\}$.

A straightforward implementation of this algorithm requires $O(n^2 \cdot t)$ operations. Indeed, there are t iterations. On each iteration the algorithm chooses a vertex of minimal weight. It takes $O(|E|)$ operations to recompute the degrees, and $O(n)$ operations per vertex to compute its weight, summing up to an overall $O(n^2 + |E|) = O(n^2)$ operations per iteration. ∎

Corollary 4.35 *Let G be a graph of order n and size $nd/2$, i.e., average degree d. If $1 \leq t \leq s$ and*

$$n\binom{d}{t} > (s-1)\binom{n}{t} , \tag{23}$$

then G contains a $K_{s,t}$ subgraph. Furthermore, the algorithm described in the proof of Theorem 4.34 (starting with $W = V$) finds a $K_{s,t}$ subgraph.

Proof: Let G have degree sequence $(d_i)_1^n$. Then by the convexity of the binomial coefficient,

$$\sum_{i=1}^{n} \binom{d_i}{t} \geq n\binom{d}{t} > (s-1)\binom{n}{t} .$$

Hence, the result follows from Theorem 4.34. ∎

Remark: In applying Corollary 4.35, we should always assume that $s \geq t$ since if (refdegcond3) holds for $s \leq t$, then it also holds when s and t are interchanged.

Corollary 4.36 *Let G be a bipartite graph with bipartition (W, U), where $|U| = n$. If*

$$\sum_{y \in W} \binom{d(y)}{t} > (s-1)\binom{n}{t} ,$$

$\sum_{y \in W} \binom{d(y)}{t} > (s-1)\binom{n}{t}$, *then G contains a $K_{s,t}$ subgraph, with s vertices in W and t in U.*

Next corollary is a constructive analogue of Lemma 4.27.

Corollary 4.37 *There is an algorithm that given an n-vertex graph $G = (V, E)$ computes an isomorphic to $K_{a,a}$ subgraph of G with $a = \Omega(\frac{\log n}{\log n^2/|E|})$ in $O(n^2 \cdot \frac{\log n}{\log n^2/|E|})$ time.*

Next theorem addresses the question of constructibility of sparse diSteiner 1-preservers for arbitrary graphs.

Theorem 4.38 *For every n-vertex (di)graph, a diSteiner 1-preserver with $O(n^2/\log n)$ arcs of weight 1 or 1/2 can be constructed in $O(n^4 \frac{(\log\log n)^2}{\log n})$ time.*

Proof: To construct a diSteiner 1-preserver with at most $O(n^2/\log n)$ arcs for an arbitrary (di)graph, one needs to invoke the procedure of extracting $K_{a,a}$ at most $O(n^2 \log\log n/\log^2 n)$ times. Indeed, in a graph with $m = \Omega(n^2/\log n)$ edges, $a = \Omega(\log n/\log(n^2/m)) = \Omega(\log n/\log\log n)$, and so a single extraction of $K_{a,a}$ results in eliminating $\Omega(\log^2 n/(\log\log n)^2)$ edges from the graph. As we start with $O(n^2)$ edges, after $O(\frac{n^2(\log\log n)^2}{\log^2 n})$ extractions, the number of edges left in the graph is $O(n^2/\log n)$. By Corollary 4.37, each extraction can be completed in $O(n^2 \cdot \log n)$ time, and so, the assertion of the theorem follows. ∎

We remark that any improvement of a factor of $\Omega(n)$ of the running time in Theorem 4.38 to $o(|E| \cdot n)$ would have some interesting applications to efficient computation of distances in dense graphs (by computing their diSteiner 1-preserver, and performing distance computations on the 1-preserver, assuming that the later is sparser than the original graph).

Next, observe that a polynomial time algorithm for constructing diSteiner D-preserver for an arbitrary (di)graph can be obtained by composing the results of Theorems 4.32 and 4.38.

Corollary 4.39 *For any n-vertex (di)graph $G = (V, E)$ and any $D = 1, 2, \ldots$, a diSteiner D-preserver with $O(n^2/\log n)$ arcs of weight 1 or 1/2 can be constructed in $O(|E|^3 \cdot n)$ time.*

4.3.3 (D, g)-Preservers

Next, we present an upper bound on $\bar{f}(D, g, n)$, that is, the size of the n-vertex extremal graph whose only (D, g)-preserver is the graph itself.

Recall that our upper bound on $f(D, n)$, that is, the minimal value such that any n-vertex graph has a D-preserver with at most $f(D, n)$ edges, was derived through the analysis of the size of the extremal graph G whose only subgraph D-preserver is G itself, i.e., $\bar{f}(D, n)$. This was possible due to the duality $f(D, n) = \bar{f}(D, n)$ (Lemma 4.3). In the case of (D, g)-preservers we are not aware of any upper bound on $f(D, g, n)$ in terms of $\bar{f}(D, g, n)$. However, we believe that the bounds on $\bar{f}(D, g, n)$ are of independent interest, and may also serve as a first step towards a better understanding the behavior of $f(D, g, n)$.

The following observation can be derived from the definition of (D, g)-preserver.

Lemma 4.40 *Every graph $G = (V, E)$ whose only (D, g)-preserver is G itself satisfies $girth(G) \geq g + 2$.*

Proof: Suppose for contradiction that $girth(G) \leq g + 1$.

Then there exists an edge $e = (u, w)$ such that $d_{G_e}(u, w) \leq g$. Since $G_e$is not a (D, g)-preserver of G there exists a pair of vertices $x, y \in V$ such that $d_G(x, y) \geq D$, and

$$d_{G_e}(x, y) \geq d_G(x, y) + g \ . \tag{24}$$

Let P be one of the shortest paths from x to y in G. Obviously, the edge e belongs to P. I.e., without loss of generality $P = (x = v_0, \ldots, v_t = u, v_{t+1} = w, \ldots, v_s = y)$, for $|P| = s$, $t = 0, 1, \ldots, s - 1$. Let P_1 be one of the shortest paths from u to w in G_e. Note that $|P_1| = d_{G_e}(u, w) \leq g$. Let $P_{x,u}$ denote the path $(x = v_0, v_1, \ldots, v_t = u)$, and $P_{w,y}$ denote the path $(v_{t+1} = w, v_{t+2}, \ldots, v_s = y)$.

Consider the walk $P_2 = P_{x,u} \cdot P_1 \cdot P_{w,y}$. Also,
$|P_2| = |P_{x,u}| + |P_1| + |P_{w,y}| \leq t + g + s - (t+1) = s + g - 1 = |P| + g - 1 = d_G(x, y) + g - 1$.
Note that $P_2 \subseteq E \setminus \{e\}$ is a path between x and y. Thus, $d_G(x, y) + g - 1 \geq |P_2| \geq d_{G_e}(x, y)$. However, this contradicts (24). ∎

Recall that for any integer $r \geq 3$, any n-vertex graph $G = (V, E)$ has at most $(n^{1+1/r-2} + n)$ edges (cf. [19], p. 166). Therefore, Lemma 4.40 implies that $\bar{f}(D, g, n) \leq n^{1+2/g} + n$. We next establish another upper bound on $\bar{f}(D, g, n)$, which is tighter whenever $D = \Omega(\sqrt{n})$.

Theorem 4.41 *For $D, g = 2, 3, \ldots$ and n sufficiently large, $\bar{f}(D, g, n) \leq 4n^{1+1/\lfloor g/4 \rfloor}/D^{1/\lfloor g/4 \rfloor}$.*

Proof: For every edge $e = (u, w) \in E$, let $P(e)$ be one of the shortest paths between $endpoint(P(e), e, u)$ and $endpoint(P(e), e, w)$ in G such that
$d_G(endpoint(P(e), e, u), endpoint(P(e), e, w)) \geq D$, but
$d_{G_e}(endpoint(P(e), e, u), endpoint(P(e), e, w)) > d_G(endpoint(P(e), e, u), endpoint(P(e), e, w)) + g$.

Note that $|subpath(P(e), e, u)| + |subpath(P(e), e, w)| \geq D - 1$.

Let $long\_subpath(P(e), e = (u, w))$ denote the longer path among $subpath(P(e), e, u)$ and $subpath(P(e), e, w)$ (if they are equal choose one of them arbitrarily).

Note that for any edge $e \in E$,

$$|long\_subpath(P(e), e)| \geq \lceil (D-1)/2 \rceil \geq D/2 - 1 \ . \qquad (25)$$

For a vertex $v \in V$, let

$$S(v) \ = \ \{e = (v, z) \in E \mid long\_subpath(P(e), e) = subpath(P(e), e, z)\} \ .$$

Consider some vertex $u \in \hat{\Gamma}_{\lfloor g/4 \rfloor - 1}(v, G)$. Let
$S(u, v) = \{e = (u, z) \in E \mid d_G(v, z) = d_G(v, u) + 1, long\_subpath(P(e), e) = subpath(P(e), e, z)\}$.

Note that $S(v) = S(v, v)$. Note also that since $girth(G) \geq g + 2$, and $d_G(v, u) \leq \lfloor g/4 \rfloor - 1$, it follows that for any edge $(u, z) \in S(u, v)$, the only shortest path from v to z in G passes through u.

Let $\hat{S}(v)$ denote the set $\bigcup_{u \in \hat{\Gamma}_{\lfloor g/4 \rfloor}(v,G)} S(u, v)$. Let $\hat{P}(v)$ denote the set

$$\hat{P}(v) \ = \ \{long\_subpath(P(e), e) \mid e \in \hat{S}(v)\} \ . \qquad (26)$$

Next, we argue that for any two paths $P_1, P_2 \in \hat{P}(v)$, $V(P_1) \cap V(P_2) = \emptyset$.

Denote $x_1 = endpoint(P_1, (u_1, z_1), u_1)$, $\quad x_2 = endpoint(P_2, (u_2, z_2), u_2)$,
$y_1 = endpoint(P_1, (u_1, z_1), z_1)$, $\quad y_2 = endpoint(P_2, (u_2, z_2), z_2)$.

Suppose for contradiction that there exists a vertex w such that $w \in V(P_1) \cap V(P_2)$.

Denote the segments of P_1 (resp., P_2) from x_1 (resp., x_2) to u_1 (resp., u_2), from u_1 (resp., u_2) to w, and from w to y_1 (resp., y_2), by P_1', P_1'' and P_1''' (resp., P_2', P_2'' and P_2'''), respectively.

Next, we show that

$$d_G(u_2, w) - (g/2 - 2) \leq d_G(u_1, w) \leq d_G(u_2, w) + (g/2 - 2) \ . \qquad (27)$$

Indeed, suppose for contradiction that $d_G(u_1, w) < d_G(u_2, w) - (g/2 - 2)$ (the case of $d_G(u_1, w) > d_G(u_2, w) + (g/2 - 2)$ is symmetrical).

Thus,

$$d_G(u_1, w) + (g/2 - 2) < d_G(u_2, w) \ . \qquad (28)$$

Then consider the path $P_{u_2,w} = P_{u_2,v} \cdot P_{v,u_1} \cdot P_1''$, where $P_{u_2,v}$ is the shortest path from u_2 to v in G, and P_{v,u_1} is the shortest path from v to u_1 in G.

Note that

$$\begin{aligned} |P_{u_2,w}| \ &= \ |P_{u_2,v}| + |P_{v,u_1}| + |P_1''| \\ &\leq \ 2(g/4 - 1) + d_G(u_1, w) \ = \ d_G(u_1, w) + g/2 - 2 \ < \ d_G(u_2, w) \end{aligned}$$

(the last inequality is by (28)).

This is a contradiction, since $P_{u_2,w}$ is a path from u_2 to w. Hence, (27) follows.

Note that $P_{u_2,v}, P_{v,u_1} \subseteq E \setminus \{e_1\}$. Consider the path $P_{12} = P_1' \cdot P_{u_1,v} \cdot P_{v,u_2} \cdot P_2'' \cdot P_1'''$. Note that P_{12} is a path between x_1 and y_1 in G_{e_1}. Since G satisfies the large-error property,

$$\begin{aligned} |P_{12}| \ &= \ |P_1'| + |P_{u_1,v}| + |P_{v,u_2}| + |P_2''| + |P_1'''| \\ &\geq \ d_{G_{e_1}}(x_1, y_1) \ \geq \ d_G(x_1, y_1) + g \ = \ |P_1'| + |P_1''| + |P_1'''| + g \ . \end{aligned}$$

I.e., $|P_{u_1,v}| + |P_{v,u_2}| + |P_2''| \geq |P_1''| + g$.

Recall that $|P_{u_1,v}| + |P_{v,u_2}| \le g/2 - 2$. Thus,

$$|P''_2| + (g/2 - 2) \ge |P_{u_1,v}| + |P_{v,u_2}| + |P''_2| \ge |P''_1| + g \ .$$

I.e., $|P''_2| \ge |P''_1| + (g/2+2)$. In other words, $d_G(u_2, w) \ge d_G(u_1, w) + (g/2+2)$, contradicting (27).

Thus, $V(P_1) \cap V(P_2) = \emptyset$. I.e., the set $\hat{P}(v)$, defined by (26), consists of vertex-disjoint paths.

Thus, for any vertex $v \in V$,

$$\sum_{e \in \hat{S}(v)} |V(long\_subpath(P(e), e))| \le n \ .$$

Hence,

$$\sum_{v \in V} \sum_{e \in \hat{S}(v)} |V(long\_subpath(P(e), e))| \le n^2 \ .$$

Using (25) it follows that

$$\sum_{v \in V} |\hat{S}(v)| \le 2n^2/D \ . \tag{29}$$

Consider a digraph $\hat{G} = (V, \hat{E})$ with the same vertex set V as the graph G, but

$$\hat{E} \ = \ \{\langle u, w\rangle \mid (u, w) \in E, long\_subpath(P((u, w)), (u, w)) = subpath(P((u, w)), (u, w), w)\} \ .$$

In other words, every edge e of the graph G is oriented towards the endpoint w from which the subpath $subpath(P(e), e, w)$ is longer.

Observe that

$$\hat{S}(v) = \{e = (u, z) \mid \langle u, z\rangle \in \hat{E}, d_{G_e}(v, u) \le \lfloor g/4\rfloor - 1\} \ .$$

Let $a_0 = 2|E|/n$ be the average degree in G. Set $C = \lfloor a_0/4\rfloor = \lfloor |E|/2n\rfloor$. We construct a graph $G' = (V', E')$ in the following way. While there is a vertex $v \in V$ with $deg_G(v) \le C$, remove v from V and all its incident edges.

Note that at most $C \cdot n \le |E|/2$ edges are removed. I.e., $|E'| \ge |E|/2$. Also, for any vertex $v \in V'$, $deg_{G'}(v) \ge C + 1 \ge |E|/2n$. Also, $girth(G') \ge girth(G) \ge g + 2$.

Consider,

$$\hat{S}'(v) \ = \ \{e = (u, z) \in E' \mid \langle u, z\rangle \in \hat{E}, d_{G'_e}(v, u) \le \lfloor g/4\rfloor - 1\} \ .$$

Note that for any vertex $v \in V'$, $\hat{S}'(v) \subseteq \hat{S}(v)$. Hence $\sum_{v \in V'} |\hat{S}'(v)| \le 2n^2/D$.

Note that for any edge $e = (u, z) \in E'$ either $\langle u, z\rangle \in \hat{E}$ or $\langle z, u\rangle \in \hat{E}$. For any edge $e = (u, z) \in E'$, denote

$$far\_endpoint(e) = \begin{cases} u, & \langle u, z\rangle \in \hat{E}, \\ z, & \langle z, u\rangle \in \hat{E}. \end{cases}$$

Note that

$$\sum_{v \in V'} |\hat{S}'(v)| \ = \ \sum_{e \in E'} |\hat{\Gamma}_{\lfloor g/4\rfloor - 1}(far\_endpoint(e), G'_e)| \ .$$

Note that since the minimal degree in G'_e is at least $|E|/2n$, and $girth(G'_e) \ge g+2$, it follows that for any edge $e \in E'$,

$$\hat{\Gamma}_{\lfloor g/4\rfloor - 1}(far\_endpoint(e), G'_e) \ \ge \ (|E|/2n - 1)^{\lfloor g/4\rfloor - 1} \ .$$

Therefore,

$$\sum_{v\in V'} |\hat{S}'(v)| = \sum_{e\in E'} |\hat{\Gamma}_{\lfloor g/4\rfloor -1}(\mathit{far\_endpoint}(e), G'_e) \geq |E'| \cdot (|E|/2n - 1)^{\lfloor g/4\rfloor -1}$$

By Theorem 4.24, we can assume that $|E| \geq 4n$. Hence,

$$\sum_{v\in V'} |\hat{S}'(v)| \geq |E|/2 \cdot \frac{|E|^{\lfloor g/4\rfloor -1}}{n^{\lfloor g/4\rfloor -1} \cdot 4^{\lfloor g/4\rfloor -1}} \geq \frac{|E|^{\lfloor g/4\rfloor}}{n^{\lfloor g/4\rfloor -1} \cdot 2 \cdot 4^{\lfloor g/4\rfloor -1}} .$$

Hence, by (29),

$$\frac{2n^2}{D} \geq \frac{|E|^{\lfloor g/4\rfloor}}{n^{\lfloor g/4\rfloor -1} \cdot 2 \cdot 4^{\lfloor g/4\rfloor -1}} .$$

Hence, $|E|^{\lfloor g/4\rfloor} D \leq 4^{\lfloor g/4\rfloor} \cdot n^{\lfloor g/4\rfloor +1}$. Thus, $|E| \cdot D^{1/\lfloor g/4\rfloor} \leq 4 \cdot n^{1+1/\lfloor g/4\rfloor}$. ▮

5 Additive Spanners Revisited

Using Distance×Size preservation Theorem (thm. 4.26) in conjunction with Lemma 3.4, it is possible to improve the results of Corollary 3.5 concerning Steiner spanners as follows.

Theorem 5.1 *For any $n = 2, 3, \ldots$, any $\Omega(1/\log n) = \delta \leq 1/2$, for any n-vertex undirected graph $G = (V, E)$ there exists Steiner additive $O(n^{(1-\delta)(1-1/\log 1/\delta)}(\log 1/\delta)^{1-1/\log 1/\delta})$-spanner with $O(n^{1+\delta})$ edges.*

Proof: By Theorem 4.26, all the distances greater or equal than $t = n^{1-\delta}$ can be preserved by a subgraph of size $O(n^{1+\delta})$. Substituting $t = n^{1-\delta}$ to Lemma 3.4 yields the assertion of this theorem. ▮

When trying to get an analogous result concerning subgraph (versus Steiner) spanners by substituting $t = n^{1-\delta}$ and using Lemma 3.4, one obtains an upper bound of roughly $O(n^{1-\delta/2})$ on the additive error of a spanner with $O(n^{1+\delta})$ edges. This is, however, weaker than an upper bound of $O(n^{1-\delta})$ that follows directly from Theorem 3.1 in conjunction with Theorem 4.26.

Using distance preservers it is possible to improve this bound to an additive error of roughly $O(n^{1-2\delta})$ for a spanner of size $O(n^{1+\delta})$. This improvement is achieved by digging deeper into the proof of Theorem 3.1 instead of using it as a blackbox. Digging deeper into the proof of Theorem 3.1 instead of using it as a blackbox makes it possible to improve this bound to an additive error of roughly $O(n^{1-2\delta})$ for a spanner of size $O(n^{1+\delta})$. Next, we sketch the proof of this improved bound.

First, let us sketch the construction of [15]. It starts with forming a *ground partition* $\mathcal{G}$, that is a partition of the entire vertex set of the graph into disjoint subsets of small diameter, called *clusters*. Consider the supergraph $\tilde{G} = (\tilde{V}, \tilde{E})$ induced by the partition $\mathcal{G}$ of the graph G. Its vertex set $\tilde{V}$ is the set of clusters of $\mathcal{G}$, and its edgeset $\tilde{E}$ is defined as $\{(C_1, C_2) \mid d_G(C_1, C_2) = 1\}$. One of the properties of the ground partition is that $|\tilde{E}| = O(n^{1+\delta})$.

After forming the ground partition $\mathcal{G}$, BFS spanning trees of all the clusters of the ground partition $\mathcal{G}$ are inserted into the edgeset H, that is constructed through the algorithm. In addition, for any pair of neighboring clusters C_1, C_2 such that $(C_1, C_2) \in \tilde{E}$, one edge $(v_1, v_2) \in E \cap (C_1 \times C_2)$ is inserted into the edgeset H. Note that so far only $O(n^{1+\delta})$ edges were inserted into H.

Then the algorithm iteratively gets rid of small clusters by unifying them into bigger ones (later on called *superclusters*,) and interconnecting the pairs of close small clusters that cannot be unified into a supercluster. The clusters are considered close if the distance between them is at most a certain threshold. The value of this threshold determines the diameters of the superclusters that are constructed. This value must always be significantly greater than the maximal diameter of the small clusters treated on the specific iteration. Since the constructed superclusters act as small clusters on the next iteration of the algorithm, the value of the threshold increases correspondingly. Since this value is a lower bound on the additive error of the spanner constructed by the algorithm, the algorithm of [15] uses constant (but growing) values of the threshold on all the iterations. A supercluster is called *active*, if it was just formed by merging other superclusters, and is going to take part in the next iteration. All the other superclusters, that is, those that were merged into bigger ones, or those that were connected to all the nearby superclusters, become *inactive* and disappear from the execution (in particular, never become active again).

The rate of the growth of the superclusters is determined by the number of still active superclusters. That is, in the beginning, when all the (super)clusters are of size $\Omega(n^{\delta})$, there are at most $O(n^{1-\delta})$ active superclusters. Hence, every supercluster that has at least $\Omega(n^{1+\delta}/n^{1-\delta}) = \Omega(n^{2\delta})$ nearby superclusters initiates forming a bigger supercluster around it. This next-generation supercluster will be of size $\Omega(n^{3\delta})$. In the next stage there are $O(n^{1-3\delta})$ active superclusters, and every supercluster that has $\Omega(n^{1+\delta}/n^{1-3\delta}) = \Omega(n^{4\delta})$ nearby superclusters initiates forming a bigger supercluster around it. Hence, it follows that the rate of the growth of the superclusters in the algorithm of [15] is exponential in the number of iterations.

Our algorithm, unlike the one of [15], has to use distance thresholds that depend polynomially on n. This is because using constant distance thresholds yields a multiplicative error of $(1+\Omega(1))$ (specifically, $1+\epsilon$ for arbitrarily small but *constant* ϵ), which, in turn, results in increasing original distances $d_G(u, w)$ by an additive term of $\Omega(d_G(u, w))$. Loosely speaking, increasing the distance thresholds in different iterations to n^{ν} some $\nu > 0$ leads to a multiplicative error of $1+O(1/n^{h(\nu)})$ for some function $h(\nu)$, which, in turn, results in decreasing one of the terms of the additive error to $O(d_G(u, w)/n^{h(\nu)})$. Note that the distance×size preservation theorem enables to "get rid" of all the pairs of vertices u, w that are at distance $\Omega(n^{1-\delta})$ one from another. This implies a bound of $O(n^{1-\delta-h(\nu)})$ on this term of the additive error.

However, the additive error is also no smaller than the largest distance threshold (recall that the distance threshold grows from one iteration to another), as the latter determines the diameters of the superclusters that are constructed during the algorithm. Another problem with using high values of distance thresholds is that they start affecting the rate of the growth of the superclusters. To exemplify this point, suppose we use a distance threshold of n^{δ}. When having at most $n^{1-\delta}$ superclusters, each having no more than $n^{2\delta}$ nearby superclusters (located at distance no greater than the distance threshold n^{δ}), interconnecting every pair of nearby superclusters by the shortest paths will require $O(n^{1-\delta}n^{2\delta} \cdot n^{\delta}) = O(n^{1+2\delta})$ edges, which is, however, exceeds the allowed size $O(n^{1+\delta})$. (Note that this problem is not present when considering Steiner spanners, and this is exactly the reason for better upper bounds for them.) It follows that (in this example) at most $O(n^{\delta})$ clusters can be merged into a supercluster, and each supercluster of new generation will be of size $O(n^{2\delta})$ (instead of $O(n^{3\delta})$), and the rate of the growth of the superclusters becomes at most linear (instead of the exponential) in the number of iterations. This, in turn, results in increasing the number of iterations required to treat all the clusters, and, therefore, in an increase of the additive error.

Calibrating the distance threshold and other parameters of the construction to achieve the smallest additive error is the main technical problem of this section. The main result is given by the following theorem.

Theorem 5.2 *For every n-vertex graph* $G = (V, E)$ *and for every fixed* $0 < \delta < 1$ *there exists an additive* $O(2^{1/\delta} n^{(1-\delta)\frac{\lceil 1/\delta \rceil - 2}{\lceil 1/\delta \rceil - 1}})$*-spanner, constructible in polynomial time,* $H \subseteq E$ *of size* $|H| = O(n^{1+\delta}/\delta)$*.*

The formal proof of this theorem will be described in full elsewhere.

Acknowledgements

The third author is grateful to David Peleg, Mikkel Thorup, Mike Saks, Uri Zwick, Cyril Gavoile, Michael Langberg and Tadao Takaoka for helpful discussions, and to an anonymous referee for pointing out the reference [16].

References

[1] I. Althöfer, G. Das, D. Dobkin, D. Joseph and J. Soares, On sparse spanners of weighted graphs, *Discrete & Computational Geometry* **9**, (1993), 81–100.

[2] B. Awerbuch and D. Peleg, Sparse partitions, *Proc. 31st IEEE Symp. on Foundations of Computer Science*, 503–513, October 1990.

[3] B. Awerbuch and D. Peleg. Network synchronization with polylogarithmic overhead, In *Proc. 31st Symp. on Foundations of Computer Science*, pp. 514-522, October, 1990.

[4] B. Awerbuch, B. Berger, L. Cowen, D. Peleg, Near-linear time construction of sparse neighborhood covers, *SIAM Journal of Computing*, Vol. 28, **1**, 263-277, 1998.

[5] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics, vol. **184**, Springer-Verlag, New York, 1998, xiv+394 pp. (OR, Second Edition, 2002)

[6] B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics **73**, Cambridge University Press, Cambridge, 2001, xviii+498 pp.

[7] B. Bollobás, D. Coppersmith, M. L. Elkin, Sparse distance preservers and additive spanners, in *Proc. 14th Annual ACM-Siam Symp. on Discrete Algorithms*, pp. 414-423, USA, MD, Baltimore, January, 2003

[8] B. Chandra, G. Das, G. Narasimhan and J. Soares, New sparseness results on graph spanners, *Proc. 8th ACM Symp. on Computat. Geometry*, 192–201, 1992.

[9] L. P. Chew, There is a planar graph almost as good as the complete graph, *2nd Symp. on Computational Geometry*, pages 169–177, 1986.

[10] E. Cohen, Fast Algorithms for constructing t-spanners and paths of stretch t, in *Proc. 34th IEEE Symp. on Foundations of Computer Science*, IEEE, Piscataway, NJ, 1993, 648-658.

[11] E. Cohen, Polylog-time and near-linear work approximation scheme for undirected shortest paths, in *Proc. of 26th ACM Symp. on Theory of Computation*, pp. 16-26, 1994.

[12] D. P. Dobkin, S. J. Friedman and K.J. Supowit, Delaunay graphs are almost as good as complete graphs, *Proc. 31st IEEE Symp. on Foundations of Computer Science*, 1987, 20–26.

[13] D. Dor, S. Halperin, U. Zwick, All pairs almost shortest paths, *SIAM Journal on Computing*, 29 (2000), pp. 1740-1759.

[14] M. L. Elkin, Computing Almost Shortest Paths, in *Proc. 20th ACM Symp. on Principles of Distributed Computing*, pp.53-63, Newport, Rhode Island, August, 2001.

[15] M. L. Elkin and D. Peleg. $(1+\epsilon,\beta)$-Spanner Constructions for General Graphs, *Proc. 33rd ACM Symp. on Theory of Computing*, pp. 173-182, Crete, Greece, July, 2001.

[16] T. Feder and R. Motwani. Clique Partitions, Graph Compression and Speeding-up Algorithms. Special Issue for the STOC conference, Journal of Computer and System Sciences, 51 (1995):261-272.

[17] C. Gavoille, D. Peleg, S. Perennes, R. Raz. Distance labeling in graphs. *Proc. to 12th Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 210-219, Washington, DC, USA, January, 2001.

[18] R. L. Graham, B. L. Rothschield, J. H. Spencer, *Ramsey Theory*, Second Edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1990, xii+196 pp.

[19] D. Peleg, *Distributed computing: a locality-sensitive approach*, SIAM, Philadelphia, PA, 2000.

[20] D. Peleg, Proximity-preserving labeling schemes, *J. Graph Theory*, 33:167-176, 2000.

[21] D. Peleg and A. Schäffer. Graph Spanners, *Journal of Graph Theory* **13** (1989), 99-116.

[22] D. Peleg and J. D. Ullman, An optimal synchronizer for the hypercube, *SIAM J. Computing* **18**, (1989), 740–747.

[23] M. Thorup, Oct. 2001, private communication.

[24] M. Thorup, Compact Oracles for Reachability and Approximate Distances on Planar Digraphs, *Proc. 42nd Annual Symp. on Foundations of Computer Science*, Las Vegas, Nevada, Oct., 2001.

[25] M. Thorup and U. Zwick, Approximate Distance Oracles, *Proc. 33rd ACM Symp. on Theory of Computing*, pp. 183-192, Crete, Greece, July, 2001.

[26] J. D. Ullman, M. Yannakakis, High-Probability Parallel Transitive-Closure-Algorithms. SIAM J. Comput. 20(1): 100-125 (1991).