Distance-preserving Subgraphs of Interval Graphs

Kshitij Gajjar* and Jaikumar Radhakrishnan†

Tata Institute of Fundamental Research, Mumbai

August 11, 2017

Abstract

We consider the problem of finding small distance-preserving subgraphs of undirected, unweighted interval graphs with $k$ terminal vertices. We prove the following results.

1. Finding an optimal distance-preserving subgraph is NP-hard for general graphs.
2. Every interval graph admits a subgraph with $O(k)$ branching vertices that approximates pairwise terminal distances up to an additive term of +1.
3. There exists an interval graph $G_{int}$ for which the +1 approximation is necessary to obtain the $O(k)$ upper bound on the number of branching vertices. In particular, any distance-preserving subgraph of $G_{int}$ has $\Omega(k \log k)$ branching vertices.
4. Every interval graph admits a distance-preserving subgraph with $O(k \log k)$ branching vertices, i.e. the $\Omega(k \log k)$ lower bound for interval graphs is tight.
5. There exists an interval graph such that every optimal distance-preserving subgraph of it has $O(k)$ branching vertices and $\Omega(k \log k)$ branching edges, thereby providing a separation between branching vertices and branching edges.

The $O(k)$ bound for distance-approximating subgraphs follows from a naïve analysis of shortest paths in interval graphs. $G_{int}$ is constructed using bit-reversal permutation matrices. The $O(k \log k)$ bound for distance-preserving subgraphs uses a divide-and-conquer approach. Finally, the separation between branching vertices and branching edges employs Hansel’s lemma [Han64] for graph covering.

1 Introduction

We consider the following problem. Given an undirected, unweighted graph $G$ with $k$ vertices designated as terminals, our goal is to construct a small subgraph $H$ of $G$. Our notion of smallness is non-standard: we compare solutions based on the number of vertices of degree three or more. We have the following definition.

Definition 1. Given an undirected, unweighted graph $G = (V, E)$ and a set $R \subseteq V$ (the terminals), we say that a subgraph $H(V, E')$ of $G$ is distance-preserving for $(G, R)$ if for all terminals $u, v \in R$, $d_G(u, v) = d_H(u, v)$, where $d_G$ and $d_H$ denote the distances in $G$ and $H$ respectively. Let $\deg_{\geq 3}(H)$ denote the number of vertices in $H$ with degree at least three (referred to as branching vertices). Let

$$B(G, R) = \min_{H} \deg_{\geq 3}(H),$$

where $H$ ranges over all subgraphs of $G$ that are distance-preserving for $(G, R)$. For a family of graphs $\mathcal{F}$ (such as planar graphs, trees, interval graphs), let

$$B_{\mathcal{F}}(k) = \max_{G} B(G, R),$$

where $G$ ranges over all graphs in $\mathcal{F}$, and $R$ ranges over all subsets of $V(G)$ of size $k$. ◊
In this work, we obtain essentially tight upper and lower bounds on $B_I(k)$, where $I$ is the class of interval graphs. An interval graph is the intersection graph of a family of intervals on the real line. (See Definition 9 for a more detailed description.)

**Theorem 2** (Main result). Let $I$ denote the class of interval graphs.

(a) (Upper bound) $B_I(k) = O(k \log k)$.

(b) (Lower bound) There exists a constant $c$ such that for each $k$, a positive power of two, there exists an interval graph $G_{\text{int}}$ with $|R| = k$ terminals such that $B(G_{\text{int}}, R) \geq ck \log k$. This implies that $B_I(k) = \Omega(k \log k)$.

Parts (a) and (b) imply that $B_I(k) = \Theta(k \log k)$.

**Remark (i).** Part (a) is constructive. Our proof of the upper bound can be turned into an efficient algorithm that, given an interval graph $G$ on $n$ vertices, produces the required distance-preserving subgraph $H$ of $G$ in running time polynomial in $n$.

**Remark (ii).** Our interval graphs are unweighted. If we consider the family of interval graphs with non-negative weights on their edges (i.e., $I$), then using $\alpha$-distance-preserving minors introduced by Krauthgamer and Zondiner [KZ12], Cheung et al. [CGH16] introduced the notion of distance-approximating minors.

**Definition 3.** Let $G(V, E, w)$ be an undirected graph with weight function $w : E \to \mathbb{R}^{\geq 0}$ and a set of terminals $R \subseteq V$. Then, $H(V', E', w')$ with $R \subseteq V' \subseteq V$ and weight function $w' : E' \to \mathbb{R}^{\geq 0}$ is a distance-preserving minor of $G$ if: (i) $H$ is a minor of $G$, and (ii) $d_H(u, v) = d_G(u, v) \forall u, v \in R$.

Subsequent work by Krauthgamer, Nguyêñ and Zondiner [KZ12, KNZ14] implies that $B_G(k) = \Theta(k^4)$, where $G$ is the family of all undirected graphs (see Corollary 8 (b)). In this work, we prove that it is NP-hard to determine if $B(G, R) \leq m$, when given a general graph $G \in G$, a set of terminals $R \subseteq V(G)$, and a positive integer $m$. A reduction from the set cover problem is described in Theorem 6.

Following the work of Krauthgamer and Zondiner [KZ12], Cheung et al. [CGH16] introduced the notion of distance-approximating minors.

**Definition 4.** Let $G(V, E, w)$ be an undirected graph with weight function $w : E \to \mathbb{R}^{\geq 0}$ and a set of terminals $R \subseteq V$. Then, $H(V', E', w')$ with $R \subseteq V' \subseteq V$ and weight function $w' : E' \to \mathbb{R}^{\geq 0}$ is an $\alpha$-distance-approximating minor ($\alpha$-DAM) of $G$ if: (i) $H$ is a minor of $G$, and (ii) $d_H(u, v) \leq \alpha \cdot d_G(u, v) \forall u, v \in V$.

In analogy with distance-approximating minors one may ask if interval graphs admit distance-approximating subgraphs with a small number of branching vertices.

**Theorem 5.** Every interval graph $G$ with $k$ terminals admits a subgraph $H$ with $O(k)$ branching vertices such that for all terminals $u$ and $v$ of $G$

$$d_G(u, v) \leq d_H(u, v) \leq d_G(u, v) + 1.$$  

We later provide a proof of Theorem 5 (see the proof of Theorem 15).

We now elaborate on our first motivation. The following example\(^1\) illustrates the relevance of distance-preserving (-approximating) subgraphs for interval graphs.

---

\(^1\)This is not a real-life problem, though we learnt that minimizing the number of branching vertices in shipping schedules is logistically desirable.
1.2 The Shipping Problem

The port of Bandarport is a busy seaport. Apart from ships with routes originating or terminating at Bandarport, there are many ships that dock at Bandarport en route to their final destination. Thus, Bandarport can be considered a hub for many ships from all over the world.

Consider the following shipping problem. A cargo ship starts from some port $X$ and has Bandarport somewhere on its route plan. The ship needs to deliver a freight container to another port $Y$, which is not on its route plan. The container can be dropped off at Bandarport and transferred through a series of ships arriving there until it is finally picked up by a ship that is destined for port $Y$. Thus, the container is transferred from $X$ to $Y$ via some “intermediate” ships at Bandarport.

However, there is a cost associated with transferring a container from one ship to another. This is because each transfer operation requires considerable manpower and resources. Thus, the number of ship-to-ship transfers that a container undergoes should be as small as possible.

Furthermore, there is an added cost if an intermediate ship receives containers from multiple ships, or sends containers to multiple ships. This is mainly because of the bookkeeping overhead involved in maintaining which container goes to which ship. If a ship is receiving all its containers from just one ship and sending all those containers to just one other ship, then the cost associated with this transfer is zero (since a container cannot be directed to a wrong ship if there is only one option), and this cost increases as the number of to and from ships increases.

Thus, given the docking times of ships at Bandarport, and a small subset of these ships that require a transfer of containers between each other, our goal is to devise a transfer strategy that meets the following objectives.

- Minimize the number of transfers for each container.
- Minimize the number of ships that have to deal with multiple transfers.

Representing each ship’s visit to the port as an interval on the time line, this problem can be modelled using distance-preserving (approximating) subgraphs of interval graphs. In this setting, a shortest path from an earlier interval to a later interval corresponds to a valid sequence of transfers across ships that moves forward in time. The first objective corresponds to minimizing pairwise distances between terminals; the second objective corresponds to minimizing the number of branching vertices.

Let us now quantify this. Suppose that there are a total of $n$ ships that dock at the port of Bandarport. Out of these, there are $k$ ships that require a transfer of containers between each other (typically $k \ll n$). Our results for interval graphs imply the following.

1. If we must make no more than the minimum number of transfers required for each container, then there is a transfer strategy in which the number of ships that have to deal with multiple transfers is $O(k \log k)$.
2. If we are allowed to make one more than the minimum number of transfers required for each container, then there is a transfer strategy in which the number of ships that have to deal with multiple transfers is $O(k)$.
3. Neither bound can be improved, i.e. there exist scheduling configurations in which $\Omega(k \log k)$ and $\Omega(k)$ ships, respectively, have to deal with multiple transfers.

1.3 Our Techniques

The linear upper bound mentioned in Theorem 5 is easy to prove (see Theorem 15). However, if we require that distances be preserved exactly, then the problem becomes non-trivial. We now present a broad overview of the techniques involved in proving our main result.

**The Upper Bound:** We may restrict attention to interval graphs that have interval representations where the terminals are intervals of length 0 (their left and right end points are the same) and the non-terminals are intervals of length 1. It is well-known that shortest paths in interval graphs can be constructed using a simple greedy algorithm. We build a subgraph consisting of such shortest paths.

2The container cannot be left at the warehouse/storage unit of Bandarport itself beyond a certain limited period of time.
starting at different terminals and add edges to it so that all inter-terminal shortest paths become available in the subgraph. We use a divide-and-conquer strategy, repeatedly “cutting” the graph down the middle into smaller interval graphs. Then we glue the solutions to the two smaller problems together. For this, we need a key observation (which appears to be applicable specifically to interval graphs) that allows one shortest path to “hop” onto another. In this, our upper bound method is significantly different from methods used previously for other families of graphs.

The Lower Bound: We construct an interval graph and arrange its vertices on a two-dimensional grid instead of the more natural one-dimensional number line. We then show that this grid can be thought of as a matrix, in particular, the bit-reversal permutation matrix (where the ones corresponding to terminals and the zeros to non-terminals). The bit-reversal permutation matrix has seen many applications, most notably in the celebrated Cooley-Tukey algorithm for Fast Fourier Transform [CT65]. Prior to our work too, it has been used to devise lower bounds (e.g. [FL87, PD06]). Examining the routes available for shortest paths in our interval graph (constructed using the bit-reversal permutation matrix) requires (i) an analysis of common prefixes of binary sequences, and (ii) building a correspondence between branching vertices and the \( k \log k/2 \) edges of a \( (\log k) \)-dimensional Boolean hypercube.

2 Distance-preserving Subgraphs for General Graphs

In this section, we first analyze the problem of finding optimal distance-preserving subgraphs of general graphs, and then study distance-preserving subgraphs for weighted graphs (including weighted interval graphs).

2.1 Finding Optimal Distance-preserving Subgraphs of General Graphs

In this section, we show that the algorithmic task of finding an optimal distance-preserving subgraph of a general graph is \( \text{NP} \)-hard. Consider the following task.

Input: An undirected, unweighted graph \( G \), a set of terminals \( R \subseteq V(G) \), and a positive integer \( \ell \).

Output: Yes, if \( (G, R) \) admits a distance-preserving subgraph with at most \( \ell \) branching vertices; No, otherwise.

Theorem 6. The above decision problem is \( \text{NP} \)-complete.
Proof. It is easy to see that the problem is in \( \text{NP} \). To show that it is \( \text{NP} \)-hard, we reduce the set cover problem to the above problem. Consider an instance of the set cover problem on a universe \( U \) of size \( n \), and subsets \( S_1, S_2, \ldots, S_m \subseteq U \).

Using this instance of the set cover problem, we construct \( G_{\text{set}} \), a graph on \( n(m+1)+m+2 \) vertices with \( n(m+1)+2 \) terminal vertices (Figure 1). Let \( U_1, U_2, \ldots, U_{m+1} \) be \( m+1 \) copies of \( U \).

\[
U_i = \{(u, i) : u \in U\}.
\]

Let \( U = \bigcup_i U_i \). Let \( S = \{S_1, S_2, \ldots, S_m\} \). The vertex set of \( G_{\text{set}} \) is \( U \cup S \cup \{t_0, t_1\} \). The edge set of \( G_{\text{set}} \) is \( E_0 \cup E_1 \cup E_2 \), where

\[
E_0 = \{(t_0, (u, i)) : (u, i) \in U\};
E_1 = \{((u, i), S_j) : u \in S_j \in S\};
E_2 = \{(S_j, t_1) : S_j \in S\}.
\]

The set of terminals is \( U \cup \{t_0, t_1\} \). We claim that \( G_{\text{set}} \) has a distance-preserving subgraph with at most \( \ell \) non-terminal branching vertices if and only if the set cover instance has a cover of size at most \( \ell \).

The if direction is straightforward. Simply fix a set cover of size at most \( \ell \) and consider the subgraph induced by it and the terminals.

For the only if direction, suppose there is a distance-preserving subgraph \( H \) of \( G_{\text{set}} \) that has at most \( \ell \) branching vertices. Clearly, in the distance-preserving subgraph \( H \), each \( (u, i) \) and \( t_1 \) have a common neighbour. If a vertex in \( S \) has degree at most 2 in \( H \), then it can have a neighbour in at most one \( U_i \). Since there are only \( m \) vertices in \( S \) but \( m+1 \) sets \( U_i \), there is an \( i_0 \), such that each vertex of the form \((u, i_0) \in U_{i_0}\) is a neighbour of a branching vertex in \( S \). Thus, the (at most \( \ell \)) branching vertices in \( S \) form a set cover of \( U \).

\[\square\]

2.2 Weighted Graphs

In this section, we show that \( B_{G_w}(k) = \Theta(k^4) \), where \( G_w \) is the family of all undirected graphs. This also implies results for unweighted graphs and weighted interval graphs.

**Theorem 7.** If \( G_w \) is the family of all undirected, weighted graphs, then \( B_{G_w}(k) = \Theta(k^4) \).

**Proof.** Both the upper bound proof and the lower bound proof for \( B_{G_w}(k) \) follow directly from earlier work of Krauthgamer, Nguyen and Zondiner [KNZ14].

First, we prove that \( B_{G_w}(k) = O(k^4) \). In [KNZ14, Section 2.1], they show that every undirected graph on \( k \) terminals has a distance-preserving minor with at most \( O(k^4) \) vertices. They prove this by pointing out that distance-preserving minors can be constructed by first constructing distance-preserving subgraphs, and then replacing the two edges incident on a vertex of degree two by a single new edge\(^3\). The number of vertices in the resulting minor is exactly the number of branching vertices in the distance-preserving subgraph. Thus, \( B_{G_w}(k) = O(k^4) \).

Next, we prove that \( B_{G_w}(k) = \Omega(k^4) \). The weighted planar graph (on \( O(k) \) terminal vertices and \( \Omega(k^4) \) vertices in total) exhibited in [KNZ14, Section 5] has only one distance-preserving subgraph, namely the planar graph itself. Thus, \( B_{G_w}(k) = \Omega(k^4) \), where \( P_w \) is the family of all undirected, weighted planar graphs. This implies that \( B_{G_w}(k) = \Omega(k^4) \).

\[\square\]

**Corollary 8.** (Corollaries of Theorem 7).

(a) If \( G \) is the family of all undirected, unweighted graphs, then \( B_G(k) = \Theta(k^4) \).

(b) If \( I_w \) is the family of weighted interval graphs, then \( B_{I_w}(k) = \Theta(k^4) \).

**Proof.** Since \( G \) and \( I_w \) are both sub-families of \( G_w \), the \( O(k^4) \) upper bound is straightforward. We now show the lower bound for both the cases.

Proof of (a): It is easy to see that the weighted planar graph of [KNZ14, Section 5] can be made unweighted (by subdividing the edges) so that every distance-preserving subgraph has \( \Omega(k^4) \) branching vertices.

---

\(^3\)Suppose \( x \) is a degree two vertex, and \( u \) and \( v \) are its two neighbours. Then, \((u, x)\) and \((x, v)\) are deleted from the minor, \((u, v)\) is added to the minor (if it does not already exist in the minor), and \( w((u, v)) = d(u, v) \).
Proof of (b): Theorem 7 implies that there exists a weighted graph $G$ such that every distance-preserving subgraph of $G$ has $\Omega(k^4)$ branching vertices. Let $|V(G)| = n$. Add edges of infinite (or very high) weight to $G$ so that the resulting graph is $K_n$, the complete graph on $n$ vertices. Since $K_n$ is an interval graph, this completes the proof.

\section{Interval Graphs}

We work with the following definition of interval graphs.

\begin{definition}
An interval graph is an undirected graph $G(V, E, \text{left}, \text{right})$ with vertex set $V$, edge set $E$, and real-valued functions $\text{left} : V \to \mathbb{R}$ and $\text{right} : V \to \mathbb{R}$ such that:
\begin{itemize}
  \item $\text{left}(x) \leq \text{right}(x) \quad \forall x \in V$;
  \item $(u, v) \in E \iff [\text{left}(u), \text{right}(u)] \cap [\text{left}(v), \text{right}(v)] \neq \emptyset$.
\end{itemize}

We order the vertices of the interval graph according to the endpoints of their corresponding intervals. For simplicity, we assume that all the endpoints of the intervals have distinct values. Define relations $\leq$ and $\prec$ on the set of vertices $V$ as follows.

\begin{align*}
  u \leq v & \iff \text{right}(u) \leq \text{right}(v) \quad \forall u, v \in V. \\
  u \prec v & \iff \text{right}(u) < \text{right}(v) \quad \forall u, v \in V.
\end{align*}

Note that if $u \prec v$, then $u \neq v$.
\end{definition}

\subsection{Shortest Paths in Interval Graphs}

In this section, we state some basic properties of shortest paths in interval graphs. It is well-known that one method of constructing shortest paths in interval graphs is the following greedy algorithm. Suppose we need to construct a shortest path from interval $u$ to interval $v$ (assume $u \prec v$). The greedy algorithm starts at $u$. In each step it chooses the next interval that intersects the current interval and reaches farthest to the right. It stops as soon as the current interval intersects $v$. Let $P^G_{gr}(u, v)$ be the shortest path produced by this greedy algorithm between $u$ and $v$ ($u \prec v$).

We now outline some elementary facts about greedy shortest paths, more generally about shortest paths in interval graphs. All of these facts are easy to prove.

\begin{fact}
Given an interval graph $G$ and a shortest path (not necessarily a greedy shortest path) $P_G(v_1, v_r) = (v_1, v_2, \ldots, v_r)$ in $G$, if $v_1 \prec v_r$, then $v_i \prec v_{i+1}$ for each $1 \leq i < r - 1$.
\end{fact}

\begin{fact}
Given an interval graph $G$ and three vertices $x, y, z \in V(G)$, if $x \leq y \leq z$, then $d_G(x, z) \geq d_G(y, z)$.
\end{fact}

\begin{fact}
Given an interval graph $G$, a greedy shortest path $P^G_{gr}(v_1, v_r) = (v_1, v_2, \ldots, v_r)$ in $G$, and a point $a \in \mathbb{R}$, let $B_a = \{v_i \in P^G_{gr}(v_1, v_r) : \text{left}(v_i) \leq a \leq \text{right}(v_i)\}$. Then, $|B_a| \leq 2$.
\end{fact}

\begin{fact}
Given an interval graph $G$, and a shortest path (not necessarily a greedy shortest path) $P_G(v_1, v_r) = (v_1, v_2, \ldots, v_r)$ in $G$, and a vertex $x \in V(G)$, let $B_x = \{v_i \in P_G(v_1, v_r) : x \in E(G), v_i \}$. Then, $|B_2| \leq 3$.
\end{fact}

\begin{fact}
Given an interval graph $G$, a greedy shortest path $P^G_{gr}(v_1, v_r) = (v_1, v_2, \ldots, v_r)$ in $G$, and two vertices $x, y \in V(G)$ such that $\text{left}(x) < \text{left}(y)$ and $\text{right}(x) > \text{right}(y)$. Then, $y \in P^G_{gr}(v_1, v_r)$ if and only if $v_1 = y$ or $v_r = y$.
\end{fact}

We now proceed to prove the $O(k)$ upper bound for distance-approximating subgraphs (with an additive distortion of $+1$) of interval graphs.
3.2 Distance-approximating Subgraphs of Interval Graphs

In this section, we show that a simple greedy technique yields a linear-sized distance-approximating subgraph for any interval graph. Let us restate Theorem 5.

**Theorem 15.** If $I$ is the family of all interval graphs, then there exists a subgraph $H$ of $G$ such that $d_{G \geq 3}(H) = O(k)$ and for all terminals $u$ and $v$, we have $d_G(u, v) \leq d_H(u, v) \leq 1 + d_G(u, v)$.

Let $G(V, E, \text{left, right})$ be an interval graph on $k$ terminals indexed by the set $[k]$. For any two vertices $u \preceq v$ of $G$, let $P^G_{\text{GR}}(u, v) = \text{NEXT-FARthest-RIGHT}(G, u, v)$ be the greedy shortest path between $u$ and $v$. For each $1 \leq i < k$, define the tree $T_i$ as follows.

$$T_i = \bigcup_{i < j \leq k} P^G_{\text{GR}}(i, j)$$

Thus, $T_i$ is a shortest-path tree rooted at terminal 1. We are now set to define $H_1$. This is the distance-approximating subgraph of $G$.

$$H_1 = T_1 \cup \{(v, i) \in E(G) : 1 < i < k, v \in V(T_i)\}$$

Assume that $v_{\text{last}} = k$. Then, $H_1$ may alternatively be defined as follows.

$$H_1 = P^G_{\text{GR}}(1, k) \cup \{(v, i) \in E(G) : 1 < i < k, v \in V(T_1)\}$$

It is easy to check that both these definitions are equivalent. The following theorem proves that $H_1$ approximates terminal distances in $G$ up to an additive distortion of $+1$.

**Lemma 16.** $d_G(i, j) \leq d_{H_1}(i, j) \leq d_G(i, j) + 1 \quad \forall 1 \leq i < j \leq k$.

**Proof.** For $i = 1$, $d_G(1, j) = d_{H_1}(1, j)$ since $T_1 \subseteq H_1$. Also when $(i, j) \in E(G)$, it is easily verifiable that $d_{H_1}(i, j) \leq 2$.

Now suppose $i \neq 1$. We show that for any $j$ such that $i < j \leq k$ and $(i, j) \notin E(G)$, $d_{H_1}(i, j) \leq d_G(i, j) + 1$. Let $P^G_{H_1}(i, k)$ be the greedy shortest path from $i$ to $k$ in $H_1$. For integer $p \geq 1$, let $v_{\text{GR}}(i, p)$ be the $p$-th vertex on the path $P^G_{H_1}(i, k)$ (i itself being the 0-th vertex). $v_{H_1}(i, p)$ is similarly defined. Note that right$(v_{\text{GR}}(i, p)) \geq \text{right}(v_{H_1}(i, p))$, with equality occurring when $v_{\text{GR}}(i, p) = v_{H_1}(i, p)$.

Suppose $d_G(i, j) = p$. Then, $(v_{\text{GR}}(i, p - 1), j) \in E(G)$. Using Claim 17, we know that either $v_{\text{GR}}(i, p - 1) = v_{H_1}(i, p - 1)$ or $(v_{\text{GR}}(i, p - 1), v_{H_1}(i, p - 1)) \in E(G)$. In the first case, $d_G(i, j) = d_{H_1}(i, j) = p$ and we are done. In the second case, there is a path of length at most 2 from $v_{H_1}(i, p - 1)$ to $j$. Thus, $d_{H_1}(i, j) \leq (p - 1) + 2 = p + 1$. This completes the proof.

**Claim 17.** Let $v_{\text{GR}}(i, p)$ and $v_{H_1}(i, p)$ be as defined in the proof of Lemma 16. Then for all $p \geq 1$, either $v_{\text{GR}}(i, p) = v_{H_1}(i, p)$ or $(v_{\text{GR}}(i, p), v_{H_1}(i, p)) \in E(G)$.

**Proof.** We prove this claim by inducting on $p$. For $p = 1$, the claim is trivially true. Our goal is to prove that the claim is true for $p = r + 1$, assuming that the claim is true for $p = r$. Thus, our induction hypothesis is that either $v_{\text{GR}}(i, r) = v_{H_1}(i, r)$ or $(v_{\text{GR}}(i, r), v_{H_1}(i, r)) \in E(G)$. In the first case, we have $v_{H_1}(i, r + 1) = v_{H_1}(i, r + 1)$, and we are done. In the second case, assume that $(v_{\text{GR}}(i, r), v_{H_1}(i, r)) \in E(G)$. Then, we have the following.

$$v_{\text{GR}}(i, r + 1) = \text{argmax}\{\text{right}(x) \mid x \in V(G), (v_{\text{GR}}(i, r), x) \in E(G)\}$$

$$v_{H_1}(i, r + 1) = \text{argmax}\{\text{right}(x) \mid x \in V(H_1), (v_{H_1}(i, r), x) \in E(H_1)\}$$

If $v_{\text{GR}}(i, r + 1) = v_{H_1}(i, r + 1)$, then we are done. Otherwise, let $v_{H_1}(i, r + 1) < \text{right}(v_{H_1}(i, r)) < \text{right}(v_{\text{GR}}(i, r)) \leq \text{right}(v_{H_1}(i, r + 1))$. Thus, the point right$(v_{\text{GR}}(i, r))$ is present in the interval corresponding to $v_{H_1}(i, r + 1)$ as well as in the interval corresponding to $v_{\text{GR}}(i, r + 1)$, which implies that $(v_{\text{GR}}(i, r + 1), v_{H_1}(i, r + 1)) \in E(G)$. This completes the proof of the claim.

Thus, $H_1$ approximates terminal distances in $G$ up to an additive term of $+1$. We now prove that the number of branching vertices in $H_1$ is linear in $k$.

**Lemma 18.** $H_1$ has $O(k)$ branching vertices.
Proof. For \( i \in [k] \), let \( B(1)_i = \{ v \in P^G_G(1, k) \mid (i, v) \in E(G) \} \). Then by Fact 13, \( |B(1)_i| \leq 3 \). In other words, each terminal can contribute at most 3 branching vertices to \( H_1 \). Summing over all terminals,\[
\sum_{i=1}^{k} |B(1)_i| \leq 3k
\]

\( P^G_G(1, k) \) is a simple path and thus contributes no branching vertices of its own to \( H_1 \). Since \( H_1 = P^G_G(1, k) \cup \{(v, i) \in E(G) : 1 < i < k, v \in V(T_1) \} \), \( H_1 \) has at most \( 3k \) branching vertices, completing the proof.

Lemma 16 and Lemma 18 together complete the proof of Theorem 15. Finally, we prove that this upper bound is tight by providing a matching lower bound.

Lemma 19. For every positive integer \( k \), there exists an interval graph \( G^{hard}_k \) on \( k \) terminals such that if \( H_1 \) is a subgraph of \( G^{hard}_k \) and \( H_1 \) approximates distances in \( G^{hard}_k \) up to an additive distortion of \( +1 \), then \( H_1 \) has \( \Omega(k) \) branching vertices.

Proof. Let us describe the construction of \( G^{hard}_k(V, E, \text{left}, \text{right}) \). Fix \( \epsilon = 0.01 \). \( G^{hard}_k \) has \( 2k - 2 \) non-terminal vertices \( \{v_1, v_2, \ldots, v_{2k-2}\} \) and \( k \) terminals indexed by the set \( |k| \).

\[
\begin{align*}
\text{left}(v_i) &= i - \epsilon, \\
\text{right}(v_i) &= i + \epsilon + 1 \\
\text{left}(j) &= 2j - 1.5, \\
\text{right}(j) &= 2j - 0.5
\end{align*}
\]

\( \forall 1 \leq i \leq 2k - 2. \)

\( \forall j \in [k]. \)

Suppose \( H_1 \) approximates distances in \( G^{hard}_k \) up to an additive distortion of \( +1 \). For odd \( i \), \( (v_j, v_{i+1}) \in V(H_1) \) (otherwise the terminals become disconnected in \( H_1 \)). For even \( i \), define the set \( S \) as follows (let \( j = i/2 \)).

\[ S = \{(v_{2j}, v_{2j+1}) : 1 \leq j \leq k - 2, (v_{2j}, v_{2j+1}) \in E(H_1) \} \]

Thus, \( |S| \leq k - 2 \). Consider any \( (v_{2j}, v_{2j+1}) \in S \). Then either \( (v_{2j}, j+1) \in E(H_1) \) (otherwise \( j + 1 \) becomes isolated in \( H_1 \)). Since \( (v_{2j-1}, v_{2j}) \in E(H_1) \) and \( (v_{2j}, v_{2j+2}) \in E(H_1) \), either \( v_{2j} \) or \( v_{2j+1} \) must be a branching vertex in \( H_1 \). Thus, for every edge in \( S \), at least one of its end points must be a branching vertex. Using Claim 20, \( |S| \geq k - 3 \). Combined with the fact that all the edges of \( S \) are vertex disjoint, this means that \( H_1 \) has at least \( k - 3 \) branching vertices, completing the proof.

Claim 20. Let \( S \) be as defined in the proof of Lemma 19. Then \( |S| \geq k - 3 \).

Proof. Suppose \( |S| < k - 3 \). Then, there exist terminal vertices \( j_1, j_2 \) such that \( 1 \leq j_1 < j_2 \leq k - 2 \) and \( (v_{2j_1}, v_{2j_1+1}) \notin E(H_1) \). \( (v_{2j_2}, v_{2j_2+1}) \notin E(H_1) \). This means that \( d_{H_1}(v_{2j_1}, v_{2j_1+1}) = 2 = 1 + d_{G^{hard}_k}(v_{2j_1}, v_{2j_1+1}) \), and \( d_{H_1}(v_{2j_2}, v_{2j_2+1}) = 2 = 1 + d_{G^{hard}_k}(v_{2j_2}, v_{2j_2+1}) \). Thus, \( d_{H_1}(1, k) \geq 2 + d_{G^{hard}_k}(1, k) \), and \( H_1 \) does not approximate distances in \( G^{hard}_k \) up to an additive distortion of \( +1 \), which is a contradiction. This completes the proof of the claim.

Note that our proof of the \( O(k) \) upper bound naturally translates into an algorithm. In other words, given an interval graph \( G \) on \( n \) vertices, it produces a distance-approximating subgraph \( H \) of \( G \) in running time polynomial in \( n \). We now move on to distance-preserving subgraphs of interval graphs.

4 Proof of the Upper Bound

In this section, we show that any interval graph \( G \) with \( k \) terminals has a distance-preserving subgraph with \( O(k \log k) \) branching vertices, which is simply Theorem 2 (a), restated here for completeness.

Theorem 21. If \( I \) is the family of all interval graphs, then \( B_I(k) = O(k \log k) \).

The following notation will be used in our proof. Given real numbers \( a, b \in \mathbb{R} \) such that \( a \leq b \), let \( G[a, b) \) be the induced subgraph on those vertices \( v \) with \( [\text{left}(v), \text{right}(v)] \cap [a, b) \neq \emptyset \). Similarly, let \( G[a, b] \) be the induced subgraph on those vertices \( v \) of \( G \) such that \( [\text{left}(v), \text{right}(v)] \cap [a, b) \neq \emptyset \).

We first prove the upper bound for a special case of interval graphs, and later show that the same upper bound holds (up to constants) for all interval graphs.
4.1 Interval Graphs with Point Terminals and Unit Non-terminals

Let $G$ be an interval graph on $k$ terminals such that all terminals in $G$ are zero-length intervals or point intervals and all non-terminals are unit intervals. Our goal is to obtain a distance-preserving subgraph $H$ of $G$ with $O(k \log k)$ branching vertices. Note that the $H$ that we obtain is not necessarily an interval graph. This is because $H$ need not be an induced subgraph of $G$.

Consider the greedy path $P_G^R(t_i, t_k)$ ($i < k$), where $t_k$ is the rightmost terminal. Our distance-preserving subgraph includes greedy paths from $t_i$ to $t_k$ for all $1 \leq i < k$. Let

$$H_0 = \bigcup_{1 \leq i < k} P_G^R(i, k).$$

Now, $H_0$ already provides for shortest paths from each terminal $t_i$ to $t_k$. In fact, it can be viewed as a shortest path tree with root $t_k$, but constructed backwards. Thus, the total number of branching vertices in $H_0$ is $O(k)$. We still need to arrange for shortest paths between other pairs of terminals $(t_i, t_j)$. The path $P_G^R(t_i, t_j)$ (for $i < j < k$) is either entirely contained in $P_G^R(t_i, t_k)$, or it follows $P_G^R(t_i, t_k)$ until it reaches a neighbour of $t_j$ and then branches off to connect to $t_j$. We can consider including all paths of the form $P_G^R(t_i, t_j)$ in $H_0$. That is, we need to link each such $t_j$ to vertices from $H_0$ so that each path $P_G^R(t_i, t_j)$ becomes available. If this is done without additional care, we might end up introducing $\Omega(k)$ additional branching vertices per terminal, and $\Omega(k^2)$ branching vertices in all, far more than we claimed.

The crucial idea for overcoming this difficulty is contained in the following lemma.

**Lemma 23.** Suppose $v \prec w$ and $d(v, w) = 1$. Let $(v, v_1, v_2, \ldots, v_{\ell})$ and $(w, w_1, w_2, \ldots, w_{\ell'})$ be greedy shortest paths starting from $v$ and $w$ respectively. Suppose $right(v_{\ell'}) < right(w_{\ell'})$. Then, $\ell \leq \ell'$.

**Proof.** Since $d(v, w) = 1$, the greedy strategy reaches at least as far in $j + 1$ steps from $v$ as it does in $j$ steps from $w$. Suppose for contradiction that $\ell > \ell'$ (that is $\ell \geq \ell' + 1$). Then, we have $right(v_{\ell'}) \leq right(v_{\ell'+1}) \leq right(w_{\ell'})$, contradicting our assumption that $right(v_{\ell'}) < right(w_{\ell'})$. 

The above lemma is crucial for the construction of our subgraph $H$. For example, suppose $t_i$ and $t_j$ both need to reach $t_r$ via a shortest path. Suppose $(w_j, t_r)$ is the last edge of $P_G^R(t_i, t_r)$ and $(w_j, t_r)$ is the last edge of $P_G^R(t_j, t_r)$. We claim that it is sufficient to include only one of these edges in $H$. If $right(w_j) < right(w_i)$, then it is enough to include the edge $(w_j, t_r)$ in $H$; as long as $t_i$ has a shortest path to $w_j$, this edge serves for shortest paths to $t_r$ from both $t_i$ and $t_j$. In the construction below, we add links to the greedy paths of $H_0$ so that we need to provide only one such edge per terminal. This idea forms the basis of the divide-and-conquer strategy which we present below.

Suppose $G$ has $2\ell$ terminals. We find a point $x$ so that both $G_{left} = G[\neg \infty, x]$ and $G_{right} = G[x, \infty]$ have $\ell$ terminals. By induction, we find distance-preserving subgraphs $H_{left}$ and $H_{right}$ of $G_{left}$ and
Lemma 23. Let us first define \( G \) with at most \( f(\ell) \) branching vertices each. The union of \( H_{\text{left}} \) and \( H_{\text{right}} \) has just \( 2f(\ell) \) branching vertices, but it does not yet guarantee shortest paths from terminals in \( H_{\text{left}} \) to terminals in \( H_{\text{right}} \). Using Lemma 23 and the discussion above, we connect each terminal \( t_j \) in \( H_{\text{right}} \) to only one of the greedy shortest paths of terminals from \( H_{\text{left}} \), and ensure that shortest paths to \( t_j \) are preserved from all terminals \( t_i \) in \( H_{\text{left}} \). This creates \( O(\ell) \) additional branching vertices and gives us a recurrence of the form

\[
f(2\ell) \leq 2f(\ell) + O(\ell),
\]

and the desired upper bound of \( O(k \log k) \). Unfortunately, there are technical difficulties in implementing the above strategy as stated. It is therefore helpful to augment \( H_0 \) by adding all greedy paths \( P^G_{\text{gr}}(t_i, t_j) \), where \( d(i, j) \leq 4 \). As a result, for each terminal \( t_i \), the first three vertices on \( P^G_{\text{gr}}(t_i, t_k) \) might become branching vertices. In all, this adds a one-time cost of \( O(k) \) branching vertices to our subgraph. We now present the argument formally.

For each \((a, b)\), let \( f(a, b) \) be the minimum number of non-terminals in a subgraph \( H^* \) of \( G[a, b] \) such that \( H_0 \cup H^* \) preserves all inter-terminal distances in \( G[a, b] \); let

\[
f(\ell) = \max_{(a, b)} f(a, b),
\]

where \((a, b)\) ranges over all pairs such that \( G[a, b] \) has at most \( \ell \) terminals. The following lemma is the basis of our induction.

Lemma 24. \((i) f(1) = 0; (ii) f(2\ell) \leq 2f(\ell) + O(\ell)\).

Proof. Part (i) is trivial. For part (ii), fix a pair \((a, b)\) such that \( G[a, b] \) has at most \( 2\ell \) terminals. If \( b - a \leq 1 \), \( H_0 \) already preserves distances between every two terminals in \( G[a, b] \). So, we may take \( H^* \) to be empty. Now assume that \( b - a > 1 \). Pick \( x \in [a, b] \) as large as possible such that (i) \( b - x \geq 1 \), and (ii) \( G[x, b] \) has at least \( \ell \) terminals.

Let \( G_{\text{left}} = G[a, x] \) and \( G_{\text{right}} = G[x, b] \). Since \( G_{\text{right}} \) has at least \( \ell \) terminals, \( G_{\text{left}} \) has at most \( \ell \) terminals. So, we obtain (by induction) a subgraph \( H_{\text{left}} \) of \( G[a, b] \) with at most \( f(\ell) \) non-terminals, such that \( H_0 \cup H_{\text{left}} \) preserves all inter-terminal distances in \( G_{\text{left}} \). If \( b - x > 1 \), then \( G_{\text{right}} \) has exactly \( \ell \) terminals, and we obtain by induction a subgraph \( H_{\text{right}} \) of \( G[a, b] \) with at most \( f(\ell) \) non-terminals such that \( H_0 \cup H_{\text{right}} \) preserves all inter-terminal distances in \( G[x, b] \). If \( b - x = 1 \), then we may take \( H_{\text{right}} \) to be empty (for \( H_0 \) already preserves inter-terminal distances in \( G[x, b] \)).

Our final subgraph \( H^* \) shall be of the form \( H_{\text{left}} \cup H_{\text{right}} \cup H_A \cup H_B \), where \( H_A \) and \( H_B \) are defined as follows. Refer to Figure 2. Let us now define \( H_A \). Let \( P_{\text{left}} \) be the set of greedy paths from the terminals in \( H_{\text{left}} \) to the terminal \( t_k \). Let \( V_A \) be the set of all non-terminal intervals of \( P_{\text{left}} \) that intersect with the interval \([x, x + 1]\). It is easy to see that any path in \( P_{\text{left}} \) contributes at most 4 non-terminals to \( V_A \). So, \( |V_A| \leq 4\ell \). Let \( H_A \) be the subgraph of \( G[a, b] \) induced by \( V_A \) and the terminals in \([x, x + 1]\).

Note that \( H_0 \cup H_{\text{left}} \cup H_{\text{right}} \cup H_A \) preserves all inter-terminal distances in \( G[a, x + 1] \) as well as all inter-terminal distances in \( G[x, b] \). In fact, it does more. For each terminal \( t_i \) in \( G[a, x] \), let \( v_i \) be the last vertex on the greedy path \( P^G_{\text{gr}}(t_i, t_k) \) that is in \( V_A \). Then, the above graph contains the greedy shortest path from every terminal \( t_j \) in \( G[a, x] \) to \( v_i \).

Now, it only remains to ensure that distances between terminals in \( G[a, x] \) and terminals in \( G[x + 1, b] \) are preserved. Let us now define \( H_B \). For each terminal \( t_j \) in \( G[x + 1, b] \), let \( v \) be the earliest interval (with respect to \( \prec \)) of \( P_{\text{left}} \) that contains \( t_j \). Then, we include the edge \((v, t_j)\) in \( H_B \). Thus, \( H_B \) contains at most one non-terminal per vertex in \( G[x + 1, b] \), that is, at most \( 2\ell \) non-terminals in all. This completes the description of \( H_A \) and \( H_B \). The final subgraph is \( H^* = H_{\text{left}} \cup H_{\text{right}} \cup H_A \cup H_B \).

Claim 25. Let \( t_i \) be a terminal in \( G[a, x] \) and \( t_r \) be a terminal in \( G[x, b] \). Then, \( H = H_0 \cup H^* \) preserves the distance between terminal \( t_i \) and \( t_r \).

Proof of Claim 25. Let \( v \) be the vertex that we attached to \( t_r \) in \( H_B \). If \( v \) is on \( P^G_{\text{gr}}(t_i, t_k) \), then it follows that \( P^G_{\text{gr}}(t_i, t_r) \) is in \( H \), and we are done. So we assume that \( v \) is not on \( P^G_{\text{gr}}(t_i, t_k) \). Then, let \( j \neq i \) be such that \( v \in P^G_{\text{gr}}(t_i, t_k) \). Then, we have paths

\[
\begin{align*}
P^G_{\text{gr}}(t_i, t_r) &= (t_i, w_1, w_2, \ldots, w_p, t_r); \\
P^G_{\text{gr}}(t_i, t_k) &= (t_i, w_1, w_2, \ldots, w_q + 1, \ldots, v_r = v, t_r),
\end{align*}
\]

10
where $v_{q+1}$ is the last vertex on $P_G^g(t_j, t_k)$ in $G[x, x+1]$, and $w_p$ is the first vertex on $P_G^g(t_l, t_r)$ such that $(w_p, v_{q+1}) \in E(G)$. From the construction of $H_A$, $(w_p, v_{q+1}) \in E(H)$. Following $v_q, (v_{q+1}, \ldots, v_r = v, t_r)$ are the subsequent vertices on $P_G^g(t_l, t_r)$. Note that: (i) $v_{q+1} \prec w_p$ (otherwise $v$ is on $P_G^g(t_l, t_k)$), (ii) $d(v_{q+1}, w_p + 1) = 1$ (both intervals contain right$(w_p)$), and (iii) right$(w_p) <$ right$(v_r)$ (since $v$ is the earliest interval of $P_{left}$ that contains $t_j$). By Lemma 23, $\ell - q - 1 \leq \ell' - p - 1$. Thus, $P_{H}(t_i, t_r)$ is no longer than $P_G^g(t_i, t_r)$.

We can now complete the proof of the upper bound. By Lemma 24, there is a subgraph $H^*$ of $G$ such that $H = H_0 \cup H^*$ preserves all inter-terminal distances in $G$, $H_0$ has $O(k)$ branching vertices and $H^*$ has $O(k \log k)$ non-terminals. It follows that $H$ has $O(k \log k)$ branching vertices.

### 4.2 Generalising to all Interval Graphs

In this section, we prove the upper bound for general interval graphs. In particular, we show that any interval graph can be reduced to the special case of the previous section. Given an interval graph $G$ on $k$ terminals, we produce a slightly modified interval graph $G'$ on $2k$ terminals, such that $G'$ has point terminals and unit non-terminals.

1. Initially, $G' = G$.
2. For each terminal $t \in V(G)$, add two vertices $t_{left}$ and $t_{right}$ to $V(G')$ such that

$$left(t_{left}) = right(t_{right}) = left(t), left(t_{right}) = right(t_{right}) = right(t).$$

Thus, $t_{left}$ and $t_{right}$ are point intervals, or intervals of length zero.
3. Designate $t_{left}$ and $t_{right}$ as terminals, and $t$ as a non-terminal. Thus, $G'$ now has $2k$ terminals.
4. For each non-terminal $x \in V(G')$, if there exists $y \in V(G')$ such that left$(y) < left(x)$ and right$(y) > right(x)$, then delete $x$ from $G'$.

Thus, after this pre-processing, the non-terminals of $G'$ have the following property: for any pair of non-terminals $u$ and $v$, left$(u) \leq left(v) \Leftrightarrow right(u) \leq right(v)$. Gardi [Gar07] shows that an interval graph that possesses this property can be equivalently represented by a unit interval graph. Thus, the induced subgraph on the non-terminals of $G'$ has a unit interval representation. It is easy to see that the (zero-length) terminals can also be placed in this unit interval representation such that it represents $G'$.

The final $G'$ is an interval graph (not necessarily a unit interval graph) such that all terminal intervals of $G'$ have length 0, and all non-terminal intervals of $G'$ have length 1. Thus, $G'$ now possesses the structure required for the construction of a distance-preserving subgraph $H'$ with $O(k \log k)$ branching vertices, as described in the proof of Theorem 21 in the previous section.

We now construct $H$ from $H'$, such that $H$ is a distance-preserving subgraph of $G$.

1. Initially, $H = H'$.
2. For each terminal $t \in V(G)$, if $t \notin V(H)$, then add $t$ to $V(H)$.
3. For each terminal $t \in V(H)$, modify the neighbourhood of $t$ in $H$ as follows.

$$nbhd_H(t) = nbhd_H(t_{left}) \cup nbhd_H(t_{right}) \cup nbhd_H(t).$$
4. For each terminal $t \in V(H)$, delete $t_{left}$ and $t_{right}$ from $H$.
5. For terminal pairs $t_1$ and $t_2$ such that $d_G(t_1, t_2) = 1$, if $(t_1, t_2) \notin E(H)$, then add $(t_1, t_2)$ to $E(H)$.

It is easy to see that the number of branching vertices in $H$ at most $O(k)$ more than the number of branching vertices in $H'$. Also, for terminal pairs $u$ and $v$ such that $d_G(u, v) > 1$ (assume $u \prec v$), a shortest path from $v_{right}$ to $v_{left}$ is a shortest path from $u$ to $v$. Thus, $H$ is a distance-preserving subgraph of $G$ with $O(k \log k)$ branching vertices, completing the proof of Theorem 21.
5 Proof of the Lower Bound

In this section, we show that there exists an interval graph $G_{\text{int}}$ such that any distance-preserving subgraph of $G_{\text{int}}$ has $\Omega(k \log k)$ branching vertices, which is simply Theorem 2 (b), restated here for completeness.

**Theorem 26.** If $\mathcal{I}$ is the family of all interval graphs, then $\mathbb{B}(\mathcal{I}) = \Omega(k \log k)$.

5.1 Preliminaries

We first set up some terminology that we use in this section. Let $k = 2^\gamma$, where $\gamma$ is a positive integer. We identify the numbers in the set $\{0, 1, \ldots, k-1\}$ with elements of $\{0, 1\}^\gamma$ using their $\gamma$-bit binary representation. We index the bits of the binary strings from left to right using integers $i = 1, 2, \ldots, \gamma$. Thus, $x[i]$ denotes the $i$-th bit of $x$ (from the left); we use $x[i, j]$ to denote the string $x[i]x[i+1] \ldots x[j]$ of length $j - i + 1$ (here $i, j$ satisfy $1 \leq i \leq j \leq \gamma$).

For a string of bits $a$, we use $\text{rev}_\gamma(a)$ to represent the reverse of $a$, that is, the binary string obtained by writing the bits of $a$ in the reverse order (e.g., $\text{rev}_\gamma(00010) = 01000$). We may arrange binary strings in a binary tree. Refer to Figure 3 for an example. The root is the empty string; the left child of a vertex $x$ is the vertex $x0$, and its right child is the vertex $x1$. In particular, the string $y$ is a descendant of the string $x$ if $y$ is obtained by concatenating $x$ with some (possibly empty) string $z$, that is, $y = xz$.

Consider the binary tree of depth $\gamma$, whose leaves correspond to elements of $\{0, 1\}^\gamma$. For distinct elements $x, y \in \{0, 1\}^\gamma$, let $\text{lca}(x, y)$ be the lowest common ancestor of $x$ and $y$ defined as follows:

$$\text{lca}(x, y) = x[1, \ell - 1] = y[1, \ell - 1],$$

where $\ell = \min \{i \in [\gamma] : x[i] \neq y[i]\}$.

For example, $\text{lca}(0100111, 0101010) = 010$. Let $[\text{lca}(x, y)]$ be the floor of $\text{lca}(x, y)$, and $[\text{lca}(x, y)]$ be the ceiling of $\text{lca}(x, y)$ defined as follows:

$$[\text{lca}(x, y)] = \text{lca}(x, y) \underline{01}^{\gamma - \ell},$$

$$[\text{lca}(x, y)] = \text{lca}(x, y) \underline{10}^{\gamma - \ell}.$$

Since $[\text{lca}(x, y)], [\text{lca}(x, y)] \in \{0, 1\}^\gamma$, we may regard $[\text{lca}(x, y)]$ and $[\text{lca}(x, y)]$ as numbers in the set $\{0, 1, \ldots, k - 1\}$. Note that $[\text{lca}(x, y)] = [\text{lca}(x, y)] - 1$, and if $x < y$, then $[\text{lca}(x, y)] \in [x, y)$ and $[\text{lca}(x, y)] \in (x, y)^4$.

Strings in $\{0, 1\}^\gamma$ can also be viewed as vertices of an $\gamma$-dimensional hypercube, with edge set

$$\mathcal{H}_\gamma = \{(x, x') : x, x' \in \{0, 1\}^\gamma \text{ and } x < x' \} \text{ and } \text{Ham}(x, x') = 1,$$

where Ham$(x, x')$ is the Hamming distance between $x$ and $x'$. Thus, if $(x, x') \in \mathcal{H}_\gamma$, then $x$ and $x'$ differ at a unique location where $x$ has a zero and $x'$ has a one.

**Claim 27.** Suppose $(x, x')$ and $(y, y')$ are distinct edges of $\mathcal{H}_\gamma$.

(a) If $\text{lca}(x, x') = \text{lca}(y, y')$, then $[\text{rev}_\gamma(x), \text{rev}_\gamma(x')] \cap [\text{rev}_\gamma(y), \text{rev}_\gamma(y')] = \emptyset$.

(b) If $\{[\text{lca}(x, x')], [\text{lca}(y, y')]\} \subseteq [x, x') \cap [y, y')$, then

$$[\text{rev}_\gamma(x), \text{rev}_\gamma(x')] \cap [\text{rev}_\gamma(y), \text{rev}_\gamma(y')] = \emptyset.$$

**Proof.** Although part (b) implies part (a), it is easier to show part (a) first, and then derive part (b) from it. For part (a), let $[\text{lca}(x, x')] = [\text{lca}(y, y')] = \ell - 1$. Let $a, b \in \{0, 1\}^{\gamma - \ell}$ be such that

$$a = x[\ell + 1, \gamma] = x'[\ell + 1, \gamma] \neq y[\ell + 1, \gamma] = y'[\ell + 1, \gamma] = b.$$

In particular, we have $a \neq b$ (implying $\text{rev}_{\gamma - \ell}(a) \neq \text{rev}_{\gamma - \ell}(b)$). Note that $\text{rev}_\gamma(a)$ represents the $\gamma - \ell$ most significant bits of $\text{rev}_\gamma(x)$ and $\text{rev}_\gamma(x')$; similarly, $\text{rev}_\gamma(b)$ represents the $\gamma - \ell$ most significant bits of $\text{rev}_\gamma(y)$ and $\text{rev}_\gamma(y')$.

If $\text{rev}_{\gamma - \ell}(a) < \text{rev}_{\gamma - \ell}(b)$ then $\text{rev}_\gamma(x') < \text{rev}_\gamma(y)$; and if $\text{rev}_{\gamma - \ell}(b) < \text{rev}_{\gamma - \ell}(a)$ then $\text{rev}_\gamma(y') < \text{rev}_\gamma(x)$. In either case, $[\text{rev}_\gamma(x), \text{rev}_\gamma(x')]$ and $[\text{rev}_\gamma(y), \text{rev}_\gamma(y')]$ are disjoint, proving part (a).

Next, consider part (b). Suppose $[\text{lca}(x, x')], [\text{lca}(y, y')] \in [x, x') \cap [y, y')$. Since every $p \in [x, x')$ is a descendant of $\text{lca}(x, x')$, we conclude that $\text{lca}(y, y')$ is a descendant of $\text{lca}(x, x')$. Similarly, $\text{lca}(x, x')$ is a descendant of $\text{lca}(y, y')$. But then $\text{lca}(x, x') = \text{lca}(y, y')$, and part (b) follows from part (a).
5.2 Manhattan Graphs

In this section, we describe a directed grid graph $G_{\text{bit}}^k$ (which we refer to as the Manhattan graph) with $3k$ terminals. We show that any distance-preserving subgraph of $G_{\text{bit}}^k$ has $\Omega(k \log k)$ branching vertices. The graph has $k^2 + 2k$ vertices arranged in a square grid. The vertices and edges of $G_{\text{bit}}^k$ are defined as follows. (Figure 4 makes this definition easier to understand.)

1. $V(G_{\text{bit}}^k) = \{0, 1, 2, \ldots, k-1\} \times \{-1, 0, 1, \ldots, k\}$.

2. There are three kinds of edges: horizontal, upward and downward; the edge set is given by $E(G_{\text{bit}}^k) = E_{\text{hor}} \cup E_{\text{up}} \cup E_{\text{down}}$, where
   
   - $E_{\text{hor}} = \{((i, j), (i, j + 1)) : i = 0, 1, \ldots, k-1 \text{ and } j = -1, 0, \ldots, k-1\}$;
   - $E_{\text{up}} = \{((i_1, j), (i_2, j)) : 0 \leq i_2 < i_1 \leq k-1 \text{ and } j = -1, 0, \ldots, k\}$;
   - $E_{\text{down}} = \{((i_1, j), (i_2, j)) : 0 \leq i_1 < i_2 \leq k-1 \text{ and } j = -1, 0, \ldots, k\}$.

3. The edge weights are given by the function $w : E(G_{\text{bit}}^k) \rightarrow \{0, 1\}$, defined as follows: $w(e) = 1$ if $e \in E_{\text{hor}} \cup E_{\text{up}}$, and $w(e) = 0$ if $e \in E_{\text{down}}$.

The set of terminals are of the form $T = T_{\text{left}} \cup T_{\text{mid}} \cup T_{\text{right}}$, where

- $T_{\text{left}} = \{0, 1, \ldots, k-1\} \times \{-1\}$,
- $T_{\text{right}} = \{0, 1, \ldots, k-1\} \times \{k\}$,
- $T_{\text{mid}} = \{\text{rev}_{\gamma}(i), i) : i = 0, 1, \ldots, k-1\}$.

This completes the definition of $G_{\text{bit}}^k$.

Fix an optimal distance-preserving subgraph $H_{\text{bit}}^k$ of $G_{\text{bit}}^k$. We shall show that $H_{\text{bit}}^k$ has $\Omega(k \log k)$ vertices of degree at least 3.

**Lemma 28.** $V(H_{\text{bit}}^k) = V(G_{\text{bit}}^k)$ and $E_{\text{hor}} \subseteq E(H_{\text{bit}}^k)$. 

---

Figure 3: A complete binary tree of height $\gamma$ having $k = 2^\gamma$ leaves. In this example, $\gamma = 5$, $x = 01001$ and $y = 01101$. Thus, $\text{Ham}(x, y) = 1$ and $|\text{lca}(x, y)| = 2$. 

| lca($x, y$) | lca($x, y$) |
| --- | --- |
| 0 1 | 0 1 |
Lemma 28 that every non-terminal vertex in $T_{mid}$ and all the other vertices are non-terminals. Edges are named horizontal, upward and downward in the natural way.

Figure 4: The bit-reversal permutation matrix for $k = 16$. Each cell represents a vertex: the blue cells represent the terminal vertices of $T_{mid}$; all the other vertices are non-terminals. Edges are named horizontal, upward and downward in the natural way.

Proof. For any $i \in \{0, 1, \ldots, k - 1\}$, note that the unique shortest path between the terminals $(i, -1)$ and $(i, k)$ is precisely $((i, -1), (i, 0), \ldots, (i, k))$. Thus, all vertices and all horizontal edges in the $i$-th row of $G^\text{bit}_{k}$ must be part of $H^\text{bit}_{k}$.

It follows from Lemma 28 that every non-terminal vertex in $H^\text{bit}_{k}$ has degree at least two, namely the two horizontal edges incident on it.

From now on, we rely solely on the fact that $H^\text{bit}_{k}$ is distance-preserving for every pair of terminals in $T_{mid}$, i.e. we prove the stronger statement that just preserving terminal distances in $T_{mid}$ requires $\Omega(k \log k)$ branching vertices.

Order the vertices in $T_{mid}$ as $t_0, t_1, \ldots, t_{k-1}$, where $t_i = (\text{rev}_k(i), i)$. Note that these terminals appear in different rows and columns. Consider the following pairs of terminals.

$$T_{\text{twins}} = \{(t_i, t_j) : (i, j) \in H_{\gamma}\}.$$ 

For each twin $(t_i, t_j)$, fix $P(i, j)$, a path of minimum distance between $t_i$ and $t_j$ in $H^\text{bit}_{k}$. We are now set to formally define special edges.

Definition 29. Let $\text{spcl}(i, j) = ((r_{ij}, [\text{lca}(i, j)]), (r_{ij}, [\text{lca}(i, j)]))$ be an edge of $P(i, j)$, where $\text{rev}_{\gamma}(i) \leq r_{ij} \leq \text{rev}_{\gamma}(j)$. (By Lemma 30, such an edge exists.) Let $\text{spcl} = \{\text{spcl}(i, j) : (t_i, t_j) \in T_{\text{twins}}\}$.

Lemma 30. Let $(t_i, t_j) \in T_{\text{twins}}$: let $\ell = [\text{lca}(i, j)]$. Then, there is an $r_{ij} \in [\text{rev}_{\gamma}(i), \text{rev}_{\gamma}(j)]$ such that $P(i, j)$ contains the edge $((r_{ij}, \ell), (r_{ij}, \ell + 1))$.

Proof. We have $i < j$, $t_i = (\text{rev}_{\gamma}(i), i)$ and $t_j = (\text{rev}_{\gamma}(j), j)$. Also note that since $(i, j) \in H_{\gamma}$, $\text{rev}_{\gamma}(i) < \text{rev}_{\gamma}(j)$. Thus, $d(i, j) = j - i$, and the shortest path $P(t_i, t_j)$ goes from column $i$ to column $j$ and never skips a column. Since $\ell \in [i, j]$, there must be an edge in $P(i, j)$ of the form $((r_{ij}, \ell), (r_{ij}, \ell + 1))$ (say, the edge of $P(i, j)$ that leaves column $\ell$ for the last time). We claim that $r_{ij} \in [\text{rev}_{\gamma}(i), \text{rev}_{\gamma}(j)]$. For otherwise, $P(i, j)$ would contain an additional edge from $E_{\text{hor}}$. Then, apart from the $j - i$ edges from $E_{\text{hor}}$, $P(i, j)$ would contain an additional edge from $E_{\text{up}}$ of weight 1; that is, the length of $P(i, j)$ would be at least $j - i + 1$—contradicting the fact that $d(i, j) = j - i$.

Lemma 31 (Key lemma). Suppose $(t_x, t_x')$ and $(t_y, t_y')$ are distinct pairs in $T_{\text{twins}}$ such that their special edges are in the same row $r$, that is,

$$\text{spcl}(x, x') = ((r, \alpha), (r, \alpha + 1))$$

$$\text{spcl}(y, y') = ((r, \beta), (r, \beta + 1)),$$
where $\alpha = \lfloor \text{lca}(x, x') \rfloor$ and $\beta = \lfloor \text{lca}(y, y') \rfloor$.

(a) Then, $\alpha \neq \beta$. In particular, $\text{spcl}(x, x') \neq \text{spcl}(y, y')$.

(b) Suppose $\alpha < \beta$. Then, there exists an $\ell \in [\alpha + 1, \beta]$ such that $(r, \ell)$ is either a branching vertex or a terminal in $H_k^{\text{int}}$.

Proof. Part (a) follows from Claim 27 (a). Consider part (b). By our definition of special edge, $r \in [\text{rev}_r(x), \text{rev}_r(x')]$ and $r \in [\text{rev}_r(y), \text{rev}_r(y')]$. So, $[\text{rev}_r(x), \text{rev}_r(x')] \cap [\text{rev}_r(y), \text{rev}_r(y')] \neq \emptyset$, and by Claim 27 (b) (in the contrapositive) either $\alpha \notin [y, y')$ or $\beta \notin [x, x')$. If $\alpha \notin [y, y')$, $\text{spcl}(x, x')$ is not on $P(y, y')$. The first vertex in row $r$ that is part of $P(y, y')$ is in a column $\ell \in [\alpha + 1, \beta]$. Then, $(r, \ell)$ is either a branching vertex or the terminal $t_y$. On the other hand, if $\beta \notin [x, x')$, then the last vertex of $P(t_x, t_{x'})$ in row $r$ lies in a column $\ell \in [\alpha + 1, \beta]$, so $(r, \ell)$ is either a branching vertex or the terminal $t_{x'}$. \hfill \square

Corollary 32. (Corollaries of Lemma 31).

(a) $|\text{spcl}| = |T_{\text{twins}}| = k \log k/2$ (since $|T_{\text{twins}}| = |H_\gamma| = k \log k/2$).

(b) If two edges in $\text{spcl}$ fall in the same row, then there is a branching vertex or a terminal separating them.

Theorem 33. $H_k^{\text{int}}$ has $\Omega(k \log k)$ branching vertices.

Proof. For each $i \in \{0, 1, \ldots, k - 1\}$, let $\delta_i$ be the number of distinct edges in $\text{spcl}$ in row $i$. Then, by Corollary 32 (a), we have

$$\sum_{i=0}^{k-1} \delta_i = |\text{spcl}| = \left(\frac{k \log k}{2}\right).$$

Furthermore, Corollary 32 (b) implies that there are at least $\delta_i - 2$ many branching vertices of the form $(i, x)$ in $H_k^{\text{int}}$, where $0 \leq x \leq k - 1$. Thus, the total number of branching vertices in $H_k^{\text{int}}$ is at least

$$(\delta_0 - 2) + (\delta_1 - 2) + \cdots + (\delta_{k-1} - 2) = \left(\sum_{i=0}^{k-1} \delta_i\right) - 2k = \left(\frac{k \log k}{2}\right) - 2k.$$

Since this quantity is $\Omega(k \log k)$, this completes the proof. \hfill \square

5.3 Translating the Lower Bound to Interval Graphs

In this section, we present an interval graph $G_{\text{int}}$ with $O(k)$ terminals, for which every distance-preserving subgraph has $\Omega(k \log k)$ branching vertices. Our lower bound relies on the lower bound for the Manhattan graph shown in the previous section. Figure 5 can be helpful to navigate through this proof. Let us describe the interval graph. Let $\mathcal{J}$ be the set of intervals.

$$\mathcal{J} = \{[x, x + 1] : x = -1, -1 + 1/k, \ldots, -1/k, 0, \ldots, k, k + 1/k, \ldots, k + 1 - 1/k\}.$$}

Thus, we have unit intervals starting at all integral multiples of $1/k$ in the range $[-1, k + 1 - 1/k]$; in all we have $k(k + 2)$ intervals in $\mathcal{J}$. These intervals naturally define an interval graph. Furthermore, the edges of $G_{\text{int}}$ are directed as follows. Orient the edges of $G_{\text{int}}$ from an earlier interval to a later interval, i.e. $(x, x + 1), [y, y + 1])$ is a directed edge from $[x, x + 1]$ to $[y, y + 1]$ if and only if $x < y \leq x + 1$. Note that this orientation does not affect shortest paths. Any shortest path from $[i, i + 1]$ to $[j, j + 1]$ (where $i < j$) in the undirected interval graph is also a valid directed shortest path in $G_{\text{int}}$. Also, $G_{\text{int}}$ has $k^2 + 2k$ vertices, which (surprisingly?) is the number of vertices in the Manhattan graph of the previous section. In fact, the connection is deeper. Let us arrange the intervals in a two-dimensional array

$$A = \langle a_{i,j} : i = 0, \ldots, k - 1 \text{ and } j = -1, 0, \ldots, k \rangle,$$

where $a_{i,j}$ corresponds to the interval $[j + (k - 1 - i)/k, j + 1 + (k - 1 - i)/k]$. Thus, the first $k$ intervals of $\mathcal{J}$ occupy the leftmost column of the array $A$ (from bottom to top); the next $k$ intervals occupy the
Figure 5: The transformation: (i) The set of intervals $\mathcal{J}$ (represented by their starting points) is divided into groups of size $k$ each. (ii) Then, each group is placed in a column of the 2d array $A$ from bottom to top. (iii) Finally, each slanting edge (weight 1) is replaced by a downward edge (weight 0) to obtain $H_{MH}$. Note that the distance between the pair of blue vertices is 1 in all three graphs. In $\mathcal{J}$ and $A$, they are connected by a single edge of weight 1. In $H_{MH}$, the gray vertex has a weight 1 edge to the blue vertex in its adjacent row.

next column (again from bottom to top), and so on. It is easy to check that, after this arrangement, the directed edges of $G_{int}$ are of three types: horizontal, upward and slanting.

$$E_{hor}(G_{int}) = \{(a_{i,j}, a_{i,j+1}) : 0 \leq i \leq k-1 \text{ and } -1 \leq j \leq k-1\};$$

$$E_{up}(G_{int}) = \{(a_{i,j}, a_{i',j}) : 1 \leq i \leq k-1 \text{ and } 0 \leq i' < i \text{ and } -1 \leq j \leq 1 \};$$

$$E_{slant}(G_{int}) = \{(a_{i,j}, a_{i,j+1}^{'}) : 0 \leq i \leq k-2 \text{ and } i < i' \leq k-1 \text{ and } -1 \leq j \leq k-1\}.$$

Thus, $E(G_{int}) = E_{hor}(G_{int}) \cup E_{up}(G_{int}) \cup E_{slant}(G_{int})$. All edges in $E(G_{int})$ have weight 1. This 2d array can be viewed as a $k \times (k+2)$ grid, and we place terminals in this graph at the same 3$k$ locations as in the Manhattan graph. This completes the description of $G_{int}$.

Let $H_{int}$ be a distance-preserving subgraph of $G_{int}$. Note that the terminals in the first and last column ensure that all horizontal edges must be part of $H_{int}$. So, both end points of every slanting edge and every upward edge included in $H_{int}$ are branching vertices. Our proof strategy is as follows. We obtain from $H_{int}$ a distance-preserving subgraph $H_{MH}$ of the Manhattan graph with nearly the same number of branching vertices. Since $H_{MH}$ requires $\Omega(k \log k)$ branching vertices, the number of branching vertices in $G_{int}$ is $\Omega(k \log k)$.

The transformation (Figure 5): We retain all corresponding vertices and all upward and horizontal edges of $H_{int}$ in $H_{MH}$. Now, $H_{int}$ might include slanting edges of the form $(a_{i,j}, a_{i',j+1})$ (where $i < i'$), but the pair $p = ((i,j), (i',j+1)) \notin E(G_{int})$. So, we accommodate such slanting edges in $H_{MH}$ by providing a path of weight 1 between its end points. We replace each slanting edge $(a_{i,j}, a_{i',j+1})$ in $E(H_{int})$ with the 0-weight edge $((i,j), (i',j))$ (a downward edge of weight zero) in $E(H_{MH})$. Note, that the edge $((i',j), (i',j+1))$ is a horizontal edge and is already retained in $E(H_{MH})$. Thus, $H_{MH}$ has a path of total weight 1, namely $(i,j) \rightarrow (i',j) \rightarrow (i',j+1)$, that connects the end points of the pair $p$. Let $E_{hor}(H_{int})$, $E_{up}(H_{int})$ and $E_{slant}(H_{int})$ be the horizontal, upward and slanting edges of $G_{int}$ that are
part of $H_{\text{int}}$. Then, $E(H_{\text{MH}}) = E_{\text{hor}}(H_{\text{MH}}) \cup E_{\text{up}}(H_{\text{MH}}) \cup E_{\text{down}}(H_{\text{MH}})$, where

\begin{align*}
E_{\text{hor}}(H_{\text{MH}}) &= \{((i, j), (i, j+1)) : (a_{i,j}, a_{i,j+1}) \in E_{\text{hor}}(H_{\text{int}})\}; \quad (34) \\
E_{\text{up}}(H_{\text{MH}}) &= \{((i_1, j), (i_2, j)) : (a_{i_1,j}, a_{i_2,j}) \in E_{\text{up}}(H_{\text{int}})\}; \quad (35) \\
E_{\text{down}}(H_{\text{MH}}) &= \{((i_1, j), (i_2, j)) : (a_{i_1,j}, a_{i_2,j+1}) \in E_{\text{down}}(H_{\text{int}})\}. \quad (36)
\end{align*}

It is straightforward to verify that $H_{\text{MH}}$ preserves distances between all pairs of terminals in $G_{\text{int}}^k$. However, for each slanting edge we replace, we might create a new branching vertex (for example, the vertex $(i_2, j)$ created in Equation 36 might be a branching vertex in $H_{\text{MH}}$ but with no corresponding branching vertex in $H_{\text{int}}$). The number of such vertices is at most the number of slanting edges, which in turn is at most the number of branching vertices in $H_{\text{int}}$. Thus, the total number of branching vertices in $H_{\text{MH}}$ is at most twice the number of branching vertices in $H_{\text{int}}$ (plus $O(k)$ to account for downward edges in the last column). Using Theorem 33, the number of branching vertices in $H_{\text{int}}$ is $\Omega(k \log k)$, completing the proof of Theorem 26.

6 Branching Vertices versus Branching Edges

In our formulation, we count the number of branching vertices (vertices with degree $\geq 3$). It is also reasonable to consider the number of edges incident on non-terminal branching vertices (we refer to such edges as branching edges) as the measure of complexity. Our $\Omega(k \log k)$ lower bound (Theorem 26) is clearly applicable to the number of branching edges as well.

In this section, we show a separation between the number of branching vertices and the number of branching edges. In particular, we present an interval graph $G_{\text{zero}}$ with $k$ terminals, each of length zero, such that the total number of branching edges in any distance-preserving subgraph of $G_{\text{zero}}$ must be $\Omega(k \log k)$. However, $G_{\text{zero}}$ admits a distance-preserving subgraph with $O(k)$ branching vertices.

Let us now describe $G_{\text{zero}}$. The interval representation of $G_{\text{zero}}$ has $k+1$ non-terminals of unit length each, and $k$ terminals of zero length each. Let the intervals corresponding to non-terminal vertices of $G_{\text{zero}}$ be

\[\{[x, x+k] : x = -k, -k+1, \ldots, 0\},\]

Let the intervals corresponding to terminal vertices of $G_{\text{zero}}$ be

\[\{[x, x] : x = -k, -k+1, \ldots, 0, 1, \ldots, k\}\]

See Figure 6 for an instance of $G_{\text{zero}}$.

**Theorem 37.** Every distance-preserving subgraph of $G_{\text{zero}}$ has at least $\Omega(k \log k)$ branching edges and at most $O(k)$ branching vertices.

**Proof.** Since the total number of vertices in $G_{\text{zero}}$ is $O(k)$, every distance-preserving subgraph of $G_{\text{zero}}$ has $O(k)$ branching vertices. Now we prove that every distance-preserving subgraph of $G_{\text{zero}}$ has $\Omega(k \log k)$ branching edges.
Fix a distance-preserving subgraph $H$ of $G_{\text{zero}}$. Consider pairs of terminals in the set $\{t_x : x = -k, -k+1, \ldots, -1\} \times \{t_y : y = 0, 1, \ldots, k-1\}$, and restrict attention to those pairs that are at distance two in $G_{\text{zero}}$, that is, covered by a common interval in $\mathcal{I}$. Indeed, for every pair of integers $i, j$ where $0 \leq i < j \leq k$, the pair $(t_{j-k}, t_i)$ is at distance two in $G_{\text{zero}}$. Build an auxiliary graph $\mathcal{P}$ on the vertex set $\{1, 2, \ldots, k\}$, where the pair $(i, j)$ is an edge if $(t_{j-k-1}, t_i)$ is at distance two. Clearly, $\mathcal{P}$ is a complete graph on $k$ vertices. For every interval $I \in \mathcal{I}$, let $B_I$ be the subgraph of $\mathcal{P}$ with vertex set $\{1, 2, \ldots, k\}$ and edge set $E(B_I) = \{(i, j) : (t_{j-k-1}, I) \in E(H) \text{ and } (t_i, I) \in E(H)\}$.

One can verify that $B_I$ is a bipartite graph, and the number of non-isolated vertices in $B_I$ is at most the degree of the vertex $I$ in $H$. By a result of Hansel [Han64] stated below (see also [KS67, Pip77]), the total number of isolated vertices in all $B_I$ put together is at least $k \log k$. Thus, the total number of edges in $G_{\text{zero}}$ is at least $k \log k$. Since $G_{\text{zero}}$ has only $O(k)$ vertices, at most $O(k)$ of these edges of $H$ can be incident on vertices of degree at most two. It follows that $H$ has $\Omega(k \log k)$ branching edges.

\begin{lemma}[Hansel [Han64]]\text{ Let $K_n$, the complete graph on $n$-vertices. Let $B_1, B_2, \ldots, B_r$ be bipartite graphs on vertex set $\{1, 2, \ldots, n\}$, such that $\bigcup_i E(B_i) = E(K_n)$. Suppose the number of non-isolated vertices in $B_i$ is $n_i$. Then $n_1 + n_2 + \ldots + n_r \geq k \log k$.} \end{lemma}

Acknowledgments

We are grateful to Nithin Varma and Rakesh Venkat for introducing us to the problem and helping with the initial analysis of shortest paths in interval graphs, and for their comments at various stages of this work. We would also like to thank the anonymous reviewers of this paper for their helpful suggestions and comments.

References

[CCE06] Don Coppersmith and Michael Elkin. Sparse sourcewise and pairwise distance preservers. \textit{SIAM Journal on Discrete Mathematics}, 20(2):463–501, 2006.

[CGH16] Yun Kuen Cheung, Gramoz Goranci, and Monika Henzinger. Graph Minors for Preserving Terminal Distances Approximately - Lower and Upper Bounds. In 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016), volume 55 of Leibniz International Proceedings in Informatics (LIPIcs), pages 131:1–131:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[CT65] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. \textit{Mathematics of computation}, 19(90):297–301, 1965.

[DDK+17] Karl Däubel, Yann Disser, Max Klimm, Torsten Mütze, and Frieder Smolny. Distance-preserving graph contractions. \textit{CoRR}, abs/1705.04544, 2017.

[FHW80] Steven Fortune, John Hopcroft, and James Wyllie. The directed subgraph homeomorphism problem. \textit{Theoretical Computer Science}, 10(2):111–121, 1980.

[FL87] Greg N Frederickson and Nancy A Lynch. Electing a leader in a synchronous ring. \textit{Journal of the ACM (JACM)}, 34(1):98–115, 1987.

[FM95] Tomás Feder and Rajeev Motwani. Clique partitions, graph compression and speeding-up algorithms. \textit{J. Comput. System Sci.}, 51(2):261–272, 1995.

[Gar07] Frédéric Gardi. The Roberts characterization of proper and unit interval graphs. \textit{Discrete Mathematics}, 307(22):2906 – 2908, 2007.

[Gup01] Anupam Gupta. Steiner points in tree metrics don’t (really) help. In Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms, pages 220–227. Society for Industrial and Applied Mathematics, 2001.
[Han64] Georges Hansel. Nombre minimal de contacts de fermeture n ecessaires pour r aliser une fonction bool eenne sym etrique de n variables. *CR Acad. Sci. Paris*, 258(25):6037–6040, 1964.

[KKN15] Lior Kamma, Robert Krauthgamer, and Huy L Nguyën. Cutting corners cheaply, or how to remove steiner points. *SIAM Journal on Computing*, 44(4):975–995, 2015.

[KNZ14] Robert Krauthgamer, Huy Nguyën, and Tamar Zondiner. Preserving terminal distances using minors. *SIAM Journal on Discrete Mathematics*, 28(1):127–141, 2014.

[KS67] Gyula Katona and Endre Szemerédí. On a problem of graph theory. *Studia Scientiarum Mathematicarum Hungarica*, 2:23–28, 1967.

[KZ12] Robert Krauthgamer and Tamar Zondiner. Preserving terminal distances using minors. In *Automata, Languages, and Programming*, volume 7391 of *Lecture Notes in Computer Science*, pages 594–605. Springer Berlin Heidelberg, 2012.

[LR80] Andrea S LaPaugh and Ronald L Rivest. The subgraph homeomorphism problem. *Journal of Computer and System Sciences*, 20(2):133–149, 1980.

[PD06] Mihai Pátraşcu and Erik D Demaine. Logarithmic lower bounds in the cell-probe model. *SIAM Journal on Computing*, 35(4):932–963, 2006.

[Pip77] Nicholas Pippenger. An information-theoretic method in combinatorial theory. *Journal of Combinatorial Theory, Series A*, 23(1):99–104, 1977.

[PS89] David Peleg and Alejandro A. Schäffer. Graph spanners. *Journal of Graph Theory*, 13(1):99–116, 1989.