On the Complexity of Telephone Broadcasting
From Cacti to Bounded Pathwidth Graphs

Aida Aminian, Shahin Kamali, Seyed-Mohammad Seyed-Javadi, and Sumedha

York University, Toronto, Canada
{aminian, kamalis, smjavadi}@yorku.ca, sumedhab@my.yorku.ca

Abstract

In TELEPHONE BROADCASTING, the goal is to disseminate a message from a given source vertex of an input graph to all other vertices in the minimum number of rounds, where at each round, an informed vertex can send the message to at most one of its uninformed neighbors. For general graphs of n vertices, the problem is NP-complete, and the best existing algorithm has an approximation factor of $\mathcal{O}(\log n/\log\log n)$. The existence of a constant factor approximation for the general graphs is still unknown.

In this paper, we study the problem in two simple families of sparse graphs, namely, cacti and graphs of bounded pathwidth. There have been several efforts to understand the complexity of the problem in cactus graphs, mostly establishing the presence of polynomial-time solutions for restricted families of cactus graphs (e.g., [3, 5, 12, 13, 16, 17]). Despite these efforts, the complexity of the problem in arbitrary cactus graphs remained open. We settle this question by establishing the NP-completeness of telephone broadcasting in cactus graphs. For that, we show the problem is NP-complete in a simple subfamily of cactus graphs, which we call snowflake graphs. These graphs not only are cacti but also have pathwidth 2. These results establish that, despite being polynomial-time solvable in trees, the problem becomes NP-complete in very simple extensions of trees.

On the positive side, we present constant-factor approximation algorithms for the studied families of graphs, namely, an algorithm with an approximation factor of 2 for cactus graphs and an approximation factor of $\mathcal{O}(1)$ for graphs of bounded pathwidth.

1 Introduction

The TELEPHONE BROADCASTING problem involves disseminating a message from a single given source vertex to all other vertices in a network through a series of *telephone calls*. The network is often modeled as an undirected and unweighted graph of n vertices. Communication takes place in synchronous rounds. Initially, only the source is informed. During each round, any informed vertex can transmit the message to at most one of its uninformed neighbors via a "call". The goal is to minimize the number of rounds required to inform the entire network. Hedetniemi et al. [18] identifies TELEPHONE BROADCASTING as a fundamental primitive in distributed computing and communication theory, forming the basis for many advanced tasks in these fields.

Slater et al. [24] established the NP-completeness of TELEPHONE BROADCASTING. Nonetheless, efficient algorithms have been developed for specific classes of graphs. In particular, Fraigniaud and Mitjana [8] demonstrated that the problem is solvable in polynomial time for trees. There are polynomial algorithms for several other graph families; see, e.g., [4, 10, 20, 25].

In this paper, we study cactus graphs, which are graphs in which any two cycles share at most one vertex. Cacti are a natural generalization of trees and ring graphs, providing a flexible model for applications such as wireless sensor networks, particularly when tree structures are too restrictive [1].

There have been several studies for broadcasting in specific families of cactus graphs [3, 5, 12, 13, 16, 17]. For instance, for unicyclic graphs, a simple subset of cactus graphs containing exactly one cycle, the problem is solvable in linear time [12]. Similarly, [17] proved that the chain of rings, which consists of cycles connected sequentially by a single vertex, has an optimal algorithm that runs in linear time. In k-restricted cactus graphs, where each vertex belongs to at most k cycles, for a fixed constant k, Čevnik and Žerovnik [3] proposed algorithms that compute the optimal broadcast scheme in $\mathcal{O}(n)$ time. Despite all these efforts, the complexity of TELEPHONE BROADCASTING for arbitrary cactus graphs has remained open, a question that we resolve in this paper.

TELEPHONE BROADCASTING is NP-hard for general graphs [9], and even for restricted graphs. In particular, Tale recently showed that the problem remains NP-hard for graphs of pathwidth of 3 [26], a result that naturally extends to graphs with higher pathwidths. However, the complexity of the problem for graphs of pathwidth 2 has remained an open question, which we address in this paper.

Elkin and Kortsarz [6] proved that approximating TELEPHONE BROADCASTING within a factor of $3-\epsilon$ for any $\epsilon > 0$ is NP-hard. Kortsarz and Peleg [19] showed that TELEPHONE BROADCASTING in general graphs has an approximation ratio of $\mathcal{O}(\log n/\log\log n)$. It is possible that there is a constant factor approximation for general graphs. However, constant-factor approximation exists only for restricted graph classes such as unit disk graphs [23] and certain sub-families of cactus graphs such as k-cycle graphs [2].

1.1 Contribution

This paper investigates the complexity and approximation algorithms for the TELEPHONE BROADCASTING problem for cacti and graphs of bounded pathwidth. The key contributions are summarized as follows:

- We present a simple polynomial-time algorithm, named CACTUS BROADCASTER, and prove it has an approximation factor of 2 for broadcasting in cactus graphs (Theorem 3.2). Constant-factor approximations are known for certain subfamilies of cactus graphs (e.g., k-cycle graphs [2]), and our result extends this to arbitrary cactus graphs. CACTUS BROADCASTER is reminiscent of the tree algorithm of [8] and leverages the separatability of cactus graphs. In our analysis of CACTUS BROADCASTER, we use the *k-broadcasting* model of broadcasting [11] as a reference point, where each vertex can inform up to two neighbors in a single round.

- Our main contribution is to establish the NP-completeness of TELEPHONE BROADCASTING problem in "snowflake graphs", which are a subclass of cactus graphs and also have pathwidth at most 2, therefore resolving the complexity of the problem in these graph families (Theorem 4.15). For a formal definition of snowflake graphs, refer to Definition 2.3. This hardness result is achieved through a series of reductions that start from $3,4$-SAT, a variant of the satisfiability problem that is NP-complete [27].

- We show the existence of a constant-factor approximation for graphs of bounded pathwidth. For that, we show the algorithm of Elkin and Kortsarz [7] has an approximation factor of $\mathcal{O}(4^w)$ for any graph of constant pathwidth w (Theorem 5.2). Note that this result does not rely on having a path decomposition certifying a bounded pathwidth. Constant-factor approximation algorithms are known for certain families of graphs of bounded path-width such as k-path graphs, which admit a 2-approximation algorithm [14], and our result extends this to any graph of bounded pathwidth.

1.2 Paper Structure

In Section 2, we present the preliminaries. In Section 3, we propose Cactus Broadcaster, which gives a 2-approximation for cactus graphs. In Section 4, we establish the NP-completeness of the problem in snowflake graphs. In Section 5, we present a constant-factor approximation algorithm for graphs of bounded pathwidth. We conclude in Section 6.

2 Preliminaries

For a positive integer n, we use notation $[n]$ to denote $\{1, 2, \ldots, n\}$. Also, we use $[i, j]$ to refer to denote $\{i, i+1, \ldots, j\}$. For a graph G, we use $G \setminus \{v\}$ to refer to the subgraph of G induced by all vertices of G except v.

Definition 2.1 ([18])**.** *An instance (G, s) of the* Telephone Broadcasting *problem is defined by a connected, undirected, and unweighted graph $G = (V, E)$ and a vertex $s \in V$, where s is the only informed vertex. The broadcasting protocol is synchronous and occurs in discrete rounds. In each round, an informed vertex can inform at most one of its uninformed neighbors. The goal is to broadcast the message as quickly as possible so that all vertices in V get informed in the minimum number of rounds.*

A broadcast scheme describes the ordering at which each vertex informs its neighbors. One can describe a broadcast scheme S with a *broadcast tree*, which is a spanning tree of G rooted at source s; if a vertex u is informed through vertex v, then u will be a child of v in the broadcast tree. Given that the optimal broadcast scheme of trees can be computed in linear time, a broadcast tree can fully describe the broadcast scheme. We use $\boldsymbol{br}^*(G, s)$ to refer to the number of rounds in the optimal broadcast scheme.

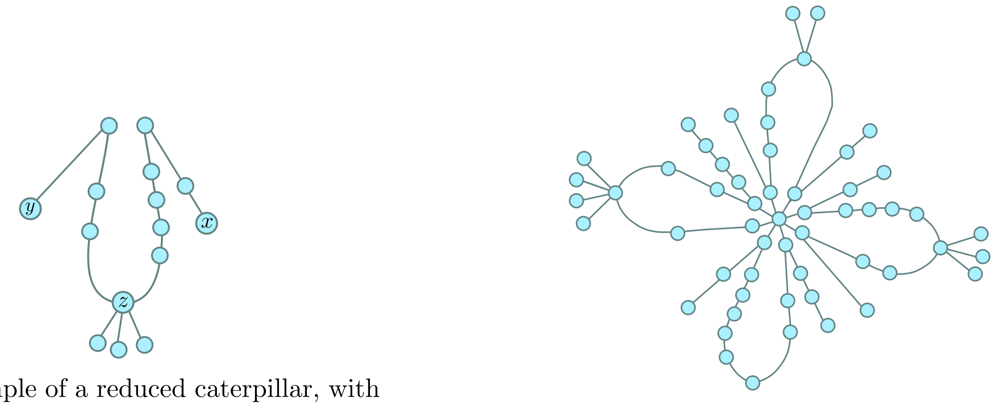
Definition 2.2. Cactus graphs *are connected graphs in which any two simple cycles have at most one vertex in common.*

Definition 2.3. *A tree T is said to be a* reduced caterpillar *if there are three special nodes x, y, and z in T such that every node in T is either located on the path between x and y or is connected to z.*

A graph G is said to be a snowflake, if and only if it has a center vertex c such that $G \setminus \{c\}$ is a set of disjoint reduced caterpillars such that c is connected to exactly two vertices in any of these caterpillars, none being special vertices.

An example of snowflake graphs and one of its corresponding reduced caterpillar components is shown in Figure 1. Informally, a snowflake graph is formed by a set of cycles that have a common center c; moreover, each cycle has a special vertex (z vertices). Any vertex in G is either i) a part of one of the cycles or ii) a part of a "dangling path" connected to a neighbor of c or iii) finds a special vertex as its sole neighbor.

Definition 2.4 ([22])**.** *A path decomposition D of a given graph $G = (V, E)$ is a sequence $\langle B_1, B_2, \ldots, B_k \rangle$, where each B_i is called a bag and contains a subset of V, such that every vertex $v \in V$ appears in at least one bag and, for every edge $(u, v) \in E$, there exists a bag B_i containing*



(a) An example of a reduced caterpillar, with the special vertices x, y, and z highlighted

(b) An example of a snowflake graph

Figure 1: An illustration of reduced caterpillar and snowflake graphs

both u and v. Furthermore, if a vertex v appears in B_i and in B_j, it must appear in any B_k where $k \in [i, j]$.

The width of the path decomposition D is the maximum cardinality of its bags minus 1. *Now, G is said to have pathwidth w if it has a path decomposition of width at most w.*

Observation 2.1. *Snowflake graphs have a pathwidth of at most 2.*

Proof. Removing the center from a snowflake graph results in a collection of disjoint caterpillar graphs, each having a pathwidth of 1 (caterpillars are graphs of pathwidth 1 [21]). A valid path decomposition for a snowflake graph can be achieved by adding the center center to all bags of path decompositions of these caterpillars. □

3 Cactus Broadcaster: A 2-Approximation for Cactus Graphs

In this section, we present our 2-approximation algorithm for cactus graphs. We use ideas from *k-broadcasting* model with parameter k [11], where a vertex can inform up to k of its neighbors in a single round via a *super call*. It is easy to see that if one can complete k-broadcasting in m rounds, then it is possible to complete broadcasting (in the classic setting) within km rounds. This can be achieved by "simulating" a super-call with up to k regular calls (see Lemma 3.1). For our algorithm, we use *k-broadcasting* with $k = 2$. In particular, we design an algorithm for *2-broadcasting* in a cactus graph G and show that if it completes within m rounds, then any broadcast scheme for classic broadcasting in G takes at least m rounds (see Lemma 3.4). Therefore, if we simulate every super call with two regular calls (in arbitrary order) using Lemma 3.1, the broadcasting completes within $2m$ rounds, and thus, we achieve an approximation factor of 2.

3.1 *k-broadcasting* Model

In the *k-broadcasting* model, an informed vertex can simultaneously inform up to k neighbors in a single round, a process we refer to as a **super call**. Creating networks that allow fast broadcasting under this model has been studied in previous work [15]. We will present a method to convert a broadcasting schema in the *k-broadcasting* model into the classic model (without super calls). The final number of rounds in the classic model will be at most k times the number of rounds in the

(a) The broadcast scheme in the *k-broadcasting* model with $k = 2$

(b) The corresponding classic broadcasting scheme

Figure 2: An illustration of Lemma 3.1. Highlighted edges show broadcast trees, and the numbers indicate the rounds at which vertices get informed.

k-broadcasting model. This model applies not only to cactus graphs but also to every arbitrary graph.

Lemma 3.1. *Let S_k be a broadcast schema for graph G in the k-broadcasting model. It is possible to convert S_k to a broadcast scheme S in linear time for graph G in the classic model such that broadcasting in S completes within k times the number of rounds as broadcasting in S_k, i.e., $\boldsymbol{br}(S) \leq k \cdot \boldsymbol{br}(S_k)$.*

Proof. Form S from S_k as follows. Suppose a vertex v informs its neighbors at rounds $(1, 2, \ldots, p)$ in S_k. In the S, v informs the same neighbors that it informs in S_k (i.e., they both have the same broadcast tree), except that each super-call in S_k is replaced by up to k regular calls in S, ensuring that if a neighbor x is informed before a neighbor y in S_k, then x gets informed before y in S as well. For example, if a vertex v first informs a neighbor a with a regular call and then neighbors b, c simultaneously with a super call right after getting informed in S_k, it will inform a at round 1 and b, c at rounds 2 and 3 (in arbitrary order) in S (see Figure 2 for an illustration).

Next, we prove that broadcasting in S completes within a factor k of S_k. We let $t_{S_k}(x)$ (respectively $t_S(x)$) denote the round at which vertex x gets informed in S_k (respectively in S). We will use an inductive argument on the time that vertices get informed to show for any vertex x, we have $T_S(x) \leq kT_{S_k}(x)$. At $t = 0$, we have $T_S(s) = T_{S_k}(s) = 0$, and the base of the induction holds. Suppose a vertex v is informed at time $t_{S_k}(v)$ in S_k, and assume v informs its neighbor x exactly d rounds after it is informed. That is, x gets informed in S_k at time $T_{S_k}(x) = T_{S_k}(v) + d$. In S, v informs x at most kd around after getting informed (because each super call is replaced by at most k regular calls). Therefore, $T_S(x) \leq T_S(v) + kd$. On the other hand, by the induction hypothesis, we have $T_S(v) \leq kT_{S_k}(v)$. We can conclude that $T_S(x) \leq kT_{S_k}(v) + kd = kT_{S_k}(x)$, which completes the induction step. □

3.2 Cactus Broadcaster Algorithm

This section presents Cactus Broadcaster, a 2-*broadcasting* algorithm that yields our 2-approximation algorithm for the Telephone Broadcasting problem in cactus graphs. Cactus Broadcaster is a mutually recursive algorithm with two main methods, namely, `single-br(`G, s`)` and `double-br(`G, s_1, s_2`)`. The `single-br(`G, s`)` method is used for 2-*broadcasting* a connected subgraph G starting from a source vertex s. The `double-br(`G, s_1, s_2`)` method is used for 2-*broadcasting* a connected subgraph G using two sources s_1 and s_2 under an assumption that both s_1 and s_2 have the message at time 0 and there is a unique path between them in G.

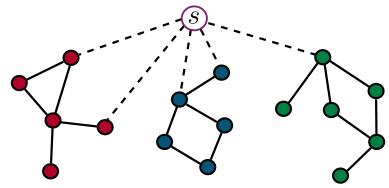


Figure 3: Two possible types for broadcasting to connected components after deleting s in `single-br`(G, s) are shown. The green vertices form a single-neighbor component. The red and blue components are double-neighbor components.

Broadcasting from a single source. We explain how `single-br`(G, s) operates. Observe that removing the source vertex s partitions G into one or more disjoint connected components. By the definition of cactus graphs, each of these connected components contains at most two neighbors of s (see Observation 3.1). We refer to a connected component with one neighbor (respectively two neighbors) of s as a *single-neighbor* (respectively *double-neighbor*) component (see Figure 3).

Observation 3.1. *Let H be a connected induced subgraph of a cactus graph G. Any vertex $v \in G$ that is not in H finds at most 2 neighbors in H.*

Proof. Since H is connected, having more than two neighbors for v would imply an edge belonging to at least two cycles, which contradicts the definition of cactus graphs. □

`single-br`(G, s) computes the broadcast time of each single-neighbor component C recursively by computing `single-br`(C, u), where u is the single neighbor of s in C. Similarly, it computes the broadcast time of each double-neighbor component C by computing `double-br`(C, u, v), where u, v are neighbors of s in C. After calculating all these broadcasting times, `single-br`(G, s) sorts these values in non-increasing order and informs the neighbors of s in this order. For that, it uses a regular call for single-neighbor components and a super-call for double-neighbor components. In other words, `single-br`(G, s), informs the component with the largest broadcasting time in the first round, the component with the second largest broadcast time next, and so on. This is reminiscent of broadcasting in trees [8], except that super-calls are used for double components. Let b_i denote the broadcast time of the i'th component C_i in this order.

The total broadcasting time for `single-br(G,s)` is then computed as $\max_i(i + b_i)$, which will be the output of `single-br(G,s)`.

Broadcasting from two sources. Next, we describe `double-br`(G, s_1, s_2). Let P be the unique path between s_1 and s_2. In any broadcast tree T of `double-br`, some vertices on P are informed through s_1 and some through s_2. Thus, exactly one edge $e \in P$ is excluded from the tree. After removing such e, the graph G will be partitioned into two disjoint connected components G_1^e and G_2^e, which are respectively informed through s_1 and s_2 (See Figure 4). The method `double-br` works by removing the edge e^* with a minimum value of $\max\{$`single-br`$(G_1^e, s_1),$ `single-br`$(G_2^e, s_2)\}$ over all $e \in P$.

Next, we explain how `double-br` finds e^*. An exhaustive approach that tries all edges $e \in P$ and calls `single-br` twice per edge may take exponential time.

To fix this, `double-br` works in two phases:

- In Phase 1, the method pre-computes the broadcast time for the connected subgraphs of G after removing vertices of P from G. By Observation 3.1, every resulting connected component after removing P has at most two neighbors in P. For any single-neighbor component C with

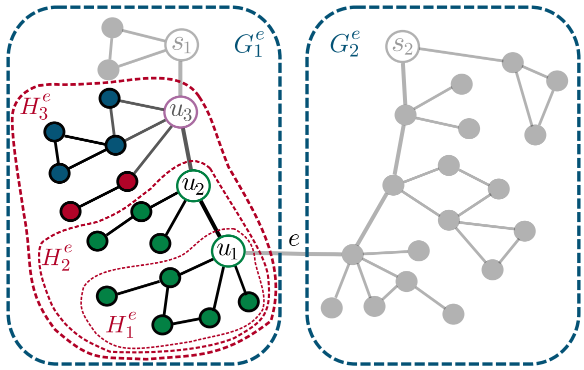


Figure 4: An illustration of Phase 2 of the `double-br` method. Different components in iteration $i = 3$ of Phase 2 (for u_3) are highlighted in different colors.

vertex s connected to P, `single-br`(C, s) is computed recursively. Similarly, for any double-neighbor component with vertices s_1, s_2 connected to P, `double-br`(C, s_1, s_2) is computed recursively.

- Next, we describe Phase 2. At each iteration of Phase 2, we fix an edge $e = (u, v) \in P$ and aim to compute `single-br`(G_1^e, s_1) and `single-br`(G_2^e, s_2), where $u \in G_1^e$ and $v \in G_2^e$. We explain how to efficiently compute `single-br`(G_1^e, s_1). Finding `single-br`(G_2^e, s_2) is done similarly. Let $\langle u(= u_1), u_2, \ldots, u_k(= s_1)\rangle$ be the sequence of vertices in the unique path from u to s_1. Let H_i^e be the connected component of G that contains u_i after removing the edge (u_i, u_{i+1}) from G_1^e. Note that $H_k^e = G_1^e$. We compute $b_i =$`single-br`(H_i^e, u_i) using the values computed in Phase 1 as follows in an iterative manner from $i = 1$ to $i = k$.

  Consider all connected components of H_i^e after removing u_i. Note that, for $i > 1$, H_{i-1}^e is one of these components. Consider the set B of broadcast times for all these components, with the neighbor(s) of u_i as the source(s) of broadcast. All these values are computed in Phase 1 except for b_{i-1}, which is computed in the previous iteration. `double-br` sorts B in non-increasing order of broadcast times and informs neighbors of u_i accordingly (similar to `single-br`). As a result, in the i'th iteration, the broadcast time b_i is computed as $\max_{j\in[B]}(j + B[j])$. At iteration $i = k$, we compute b_k, which is indeed `single-br`(G_1^e, s_1) (see Observation 3.2).

Figure 4 provides an illustration.

Theorem 3.2. Cactus Broadcaster *runs a polynomial-time for* Telephone Broadcasting *and achieves an approximation factor of* 2 *on cactus graphs.*

Proof Overview. We provide an overview of the proof before a formal argument. First, we show that Cactus Broadcaster runs in $\mathcal{O}(n^3 \log n)$. This can be established using induction on the size of the input graph, based on the recursive nature of the algorithm. Details can be found in Lemma 3.3.

Second, we show that Cactus Broadcaster has an approximation factor of 2. Intuitively, `single-br` with super-calls runs no longer than an optimal broadcast scheme without super-calls; this is because every super call only expedites informing one of the two sources of broadcasting in double-neighbor components. As mentioned earlier (Lemma 3.1), a broadcast scheme with super calls can be simulated with a scheme (without super-calls) that completes no later than twice the number of rounds. On the other hand, the selection of edge e^* by `double-br` ensures that $\max(\texttt{single-br}(G_1^{e^*}, s_1), \texttt{single-br}(G_2^{e^*}, s_2)) \leq \max(\texttt{single-br}(G_1^{e^+}, s_1), \texttt{single-br}(G_2^{e^+}, s_2))$, where e^+ is the edge that is absent in the optimal broadcast tree of G. A formal proof follows from an inductive argument on the input size (see Lemma 3.4).

Observation 3.2. *Let $\langle u_1(=u), u_2, \ldots, u_k(=s_1)\rangle$ be the vertices on the path from u_1 (an endpoint of the removed edge e) and the source s_1. Then, for any $i \in [k]$, the broadcast time b_i for (H_i^e, u_i) in `double-br` equals to `single-br` applied to H_i, that is, $b_i = \texttt{single-br}(H_i^e, u_i)$.*

Proof. We use induction on i. For $i = 1$, `double-br` simply uses the precomputed broadcast times B_i of the connected components after removing the source u_1 from H_1^e in Phase 1. It proceeds with sorting these broadcast times and setting $b_i = \max_{j\in[|B_i|]}(j + B_i[j])$. This is what `single-br` does by definition when broadcasting from u_1 in H_1. The inductive step is similar, except that when computing b_i, the value of b_{i-1} is also in B_i. This value is indeed $\texttt{single-br}(H_{i-1}^e, u_{i-1})$, by the induction hypothesis. □

Lemma 3.3. Cactus Broadcaster *terminates in $\mathcal{O}(n^3 \log n)$ time.*

Proof. To prove the lemma, we define the following functions. The functions $T_s(n)$ and $T_d(n)$ represent the maximum time required to process a subgraph of size n using `single-br` and `double-br`, respectively. Let $T(n) = \max(T_s(n), T_d(n))$.

$\texttt{single-br}(G, s)$ involves running the recursive methods with time $T(x)$ on any component of size x after removing s, along with an additional $\mathcal{O}(n \log n)$ time to sort the broadcasting times obtained from recursive calls. On the other hand, `double-br` includes precomputing the broadcast time for the subgraphs after deleting the vertices of path P between s_1 and s_2. For each of $\mathcal{O}(n)$ edges $e \in P$ that it considers for removing, we have k iterations, each computing $\texttt{single-br}(H_i, u_i)$ (for $i \in [k]$). At each iteration, we sort the broadcast times for connected components of H_i after removing u_i, which takes $\mathcal{O}(d(u_i) \log d(u_i))$, where $d(u_i)$ is the degree of u_i in G_1^e. Therefore, for a fixed edge, the total time would be $\sum_{i\in[k]} \mathcal{O}(d(u_i) \log d(u_i)) = \mathcal{O}(n \log n)$. Given that there are $\mathcal{O}(n)$ possible edges to remove, the total time complexity of `double-br` would be $\mathcal{O}(n^2 \log n)$ plus the additional time spent in Phase 1.

For each n, let $\mathcal{Q}_n$ be the set of partitions of any integer less than n to smaller integers. That is, any set of integers that sum to at most $n-1$ is a member of $\mathcal{Q}_n$. To summarize, there are some $n_0, c' \geq 0$ such that for large $n \geq n_0$, we have:

$$T_s(n) \leq c'(n \log n) + \max_{S\in\mathcal{Q}_n} \sum_{x\in S} T(x),$$
$$T_d(n) \leq c'(n^2 \log n) + \max_{S\in\mathcal{Q}_n} \sum_{x\in S} T(x).$$

Therefore,

$$\begin{aligned} T(n) =& \max(T_s(n), T_d(n)) \\ \leq& c'n^2 \log n + \max_{S\in\mathcal{Q}_n} \sum_{x\in S} T(x). \end{aligned} \tag{1}$$

We claim that $T(n) \leq cn^3 \log n$, in which $c = \max(c', T(n_0))$. To prove the claim, we use induction on n. For the base case of $n = n_0$, as $c \geq T(n_0)$, the inequality clearly holds. Next, assume $T(n') \leq cn'^3 \log(n')$ for all $n' \in [n_0, n-1]$.

$$
\begin{aligned}
T(n) &\leq c'n^2 \log n + \max_{S\in\mathcal{Q}_n} \sum_{x\in S} T(x) && \text{based on (1)} \\
&\leq c'n^2 \log n + \max_{S\in\mathcal{Q}_n} \sum_{x\in S} cx^3 \log(x) && \text{using induction hypothesis} \\
&\leq c'n^2 \log n + c(n-1)^3 \log n && \forall_{S\in\mathcal{Q}_n} \sum_{x\in S} x < n \\
&< c'n^2 \log n + (cn^3 - cn^2) \log n \\
&\leq cn^3 \log n. && c \geq c'
\end{aligned}
$$

which completes the induction step, and we can conclude $T(n) \leq cn^3 \log n$. $\square$

Lemma 3.4. *Consider an instance (G, s) of the broadcasting problem in a cactus graph G. Let $\boldsymbol{br}(G, s)$ denote the number of rounds that it takes for* Cactus Broadcaster *to complete k-broadcasting in G with $k = 2$, and $\boldsymbol{br}^*(G, s)$ denote the optimal number of rounds for broadcasting in the classic model. Then we have $\boldsymbol{br}(G, s) \leq \boldsymbol{br}^*(G, s)$.*

Proof. We use induction on the number of vertices in G to prove the claim for both methods `single-br` and `double-br`. In the base cases, the graph consists of just one (for `single-br`) or two (for `double-br`) source vertices, and the statement trivially holds.

For the induction step, we prove the claim for a graph with n vertices, assuming it holds for all graphs with $n' < n$ vertices.

First, we consider the simpler case of $\texttt{double-br(}G, s_1, s_2\texttt{)}$. Recall that the algorithm finds an edge $e^* \in P$ that minimizes

$$\max(\texttt{single-br(}G_1^{e^*}, s_1\texttt{)}, \texttt{single-br(}G_2^{e^*}, s_2\texttt{)}).$$

Assume that the optimal solution removes an edge $e^+ \in P$. Based on the induction hypothesis we know that $\boldsymbol{br}^*(G_1^{e^+}, s_1) \geq \texttt{single-br(}G_1^{e^+}, s_1\texttt{)}$ and $\boldsymbol{br}^*(G_2^{e^+}, s_2) \geq \texttt{single-br(}G_2^{e^+}, s_2\texttt{)}$. In this case, the broadcast time of the optimal solution would be

$$
\begin{aligned}
\boldsymbol{br}^*(G, s) &= \max(\boldsymbol{br}^*(G_1^{e^+}, s_1), \boldsymbol{br}^*(G_2^{e^+}, s_2)) \\
&\geq \max(\texttt{single-br(}G_1^{e^+}, s_1\texttt{)}, \texttt{single-br(}G_2^{e^+}, s_2\texttt{)}) && \text{induction hypothesis} \\
&\geq \max(\texttt{single-br(}G_1^{e^*}, s_1\texttt{)}, \texttt{single-br(}G_2^{e^*}, s_2\texttt{)}) && e^* \text{ definition} \\
&= \texttt{double-br(}G, s_1, s_2\texttt{)}.
\end{aligned}
$$

Next, we address the case of the $\texttt{single-br(}G, s\texttt{)}$. Consider the optimal broadcast scheme T^* in the classic model. For a component C_i with up to two neighbors of s, let π_i^1 and π_i^2 be the rounds at which s informs its first and its second (if exists) neighbor in C_i in T^*, respectively. Note that $\pi_i^1 < \pi_i^2$, and we let $\pi_i^2 = \infty$ if s informs just one vertex in C_i. If both neighbors of s in C_i were informed simultaneously via a super call in round π_i^1, the broadcast time would not increase compared to informing the first one at π_i^1 and the second one at round π_i^2. Hence, the following inequality holds.

$$\begin{aligned}\boldsymbol{br}^*(G,s) &= \max_i(\boldsymbol{br}^*(C_i,(\pi_i^1,\pi_i^2)))\\ &\geq \max_i(\boldsymbol{br}^*(C_i,(\pi_i^1,\pi_i^1)))\\ &= \max_i(\boldsymbol{br}^*(C_i,(0,0))) + \pi_i^1,\end{aligned}$$

where $\boldsymbol{br}^*(C_i,(\pi_i^1,\pi_i^2))$ is the optimal broadcasting time in the classic model given that the first source is informed at round π_i^1 and the second one (if exists) at round π_i^2. Recall b_i is the broadcast time for each component C_i, which is determined by applying the recursive methods on C_i. This is equivalent to informing the neighbors of s in C_i at time 0.

$$\begin{aligned}\boldsymbol{br}^*(G,s) &\geq \max_i(\boldsymbol{br}^*(C_i,(0,0))) + \pi_i^1\\ &\geq \max_i(b_i + \pi_i^1). && \boldsymbol{br}^*(C_i,(0,0)) \geq b_i \text{ (induction hypothesis)}\end{aligned}$$

`single-br` selects π_i' such that $\max_i(b_i+\pi_i')$ is minimized. This is because it sorts the broadcast times required to inform each subgraph in non-increasing order and informs the neighbor(s) of s in C_i accordingly. We can write

$$\begin{aligned}\boldsymbol{br}^*(G,s) &\geq \max_i(b_i + \pi_i^1)\\ &\geq \min_{\pi'}(\max_i(b_i + \pi_i'))\\ &= \boldsymbol{br}(G,s).\end{aligned}$$

Thus, $\boldsymbol{br}^*(G,s) \geq \boldsymbol{br}(G,s)$. Thus, the number of super rounds in both methods does not exceed the number of rounds used by the optimal broadcast scheme in the classic model. □

From the above results, we can establish the proof of Theorem 3.2 as follows.

Proof of Theorem 3.2: By Lemma 3.3, we showed that Cactus Broadcaster terminates in polynomial time. Furthermore, Lemma 3.4 establishes that the 2-*broadcasting* scheme of Cactus Broadcaster completes no later than the optimal broadcast time. Moreover, by Lemma 3.1, this 2-*broadcasting* scheme can be converted to a scheme in classic broadcasting that completes within twice the optimal broadcast time. □

The approximation factor given by Theorem 3.2 is tight. Consider a graph G formed by t triangles that all share an endpoint s. The broadcasting time of Cactus Broadcaster on (G,s) is $2t$ as its broadcast tree will be a star with $2t$ leaves, while the optimal broadcast tree completes broadcasting in $t+1$ rounds (s will have degree t). Thus, the approximation factor tends to 2 for large t.

4 Hardness Proof of Snowflake Graphs

The NP-hardness proof of Telephone Broadcasting in snowflake graphs proceeds through successive reductions. First, we reduce 3, 4-SAT, which is known to be NP-hard [27] to a new problem that we call Twin Interval Selection (Lemma 4.5), which itself reduces to another new problem Dome Selection with Prefix Restrictions (Lemma 4.12). Finally, we present a reduction from DoSePr to Telephone Broadguess, which establishes our main result (Theorem 4.15).

4.1 Hardness of Twin Interval Selection

The first step in reducing 3, 4-SAT to Telephone Broadguess is to establish a reduction from 3, 4-SAT to Twin Interval Selection (TwIS). For that, we first explain how to construct a TwIS instance from a given 3, 4-SAT instance and demonstrate the equivalence of solutions between the two problems.

Definition 4.1 ([27]). *The* 3, 4-SAT *problem is a special type of the classic* 3*-satsifiabiity (*3-SAT*) problem. The input is a boolean CNF formula $\phi = (X, C)$, where X is the set of variables and C is the set of clauses such that every clause $C_i \in C$ contains* exactly *three literals $\ell^i_k \in C_i$ for $k \in [3]$, and each variable $x \in X$ appears in at most four clauses. The decision problem asks whether there is a satisfying assignment of ϕ.*

The 3, 4-SAT problem is known to be NP-complete [27]. In what follows, we call a pair of intervals $(I_i, \overline{I_i})$ for $i \in [n]$ a *twin interval*, assuming that I_i and $\overline{I_i}$ are non-crossing and have equal lengths with endpoints in $[m]$ for some positive m. The endpoints of I_i are less than the endpoints of $\overline{I_i}$. We refer to I_i and $\overline{I}_i$ the *left* and the *right* interval of the twin, respectively.

Definition 4.2. *An instance of the* Twin Interval Selection (TwIS) *is formed by a tuple (I, r, m), where I is a set of* twin *interval pairs and r is a* restriction function *with domain $[m]$, where m is referred to the* horizon *of the instance. We have $I = \{(I_1, \overline{I_1}), \ldots, (I_n, \overline{I_n})\}$, where for each $i \in [m]$, $(I_i, \bar{I}_i)$ are non-crossing intervals with endpoints in $[m]$.*

The objective is to select exactly one of I_i and $\overline{I_i}$, while respecting the restriction imposed by the restriction function as follows. The restriction function $r : [m] \to [n]$ requires that for any $t \in [m]$, the number of selected intervals that contain t be at most $r(t)$. The decision problem asks whether there is a valid selection that satisfies these restrictions.

For a solution S and $t \in [m]$, let $\Gamma_S(t)$ be the number of selected intervals in S containing t.

Construction. Suppose we are given an instance $\phi = (X, C)$ of 3, 4-SAT. We construct an instance $\mathcal{I}_\phi$ of TwIS as follows. For each variable $x_i \in X$, form a pair of $(X_i, \overline{X_i})$ of twin intervals, each of length 7, where $X_i = [16i - 15, 16i - 8]$ and $\overline{X}_i = [16i - 7, 16i]$. We refer to X_i (respectively $\overline{X_i}$) as the *principal* interval of literal x_i (respectively $\neg x_i$). The length 7 of these intervals allows for placing up to 4 non-crossing unit intervals, each of length 1, within the principal interval; these intervals will be used for clause gadgets, as we will explain. For each clause C_j, we define three interval twin pairs $(I^j_k, \overline{I^j_k})$, each associated with one of the literals of $\ell^j_k \in C_j$ for each $k \in [3]$. Suppose ℓ^j_k is the p'th literal where its corresponding variable $x_i \in X$ appears ($p \in [4]$). If ℓ^j_k is a positive (respectively negative) literal of variable x_i, then define I^j_k as a unit interval starting at $16i - 7 + 2(p-1)$ (respectively, starting at $16i - 15 + 2(p-1)$). Intuitively, I^j_k is defined as one of the four-unit intervals within the principal interval of $\neg x$. Meanwhile, $\overline{I^j_k}$ is $[16n + 2j, 16n + 2j + 1]$ for all $k \in [3]$. Finally, to define the restriction function, we let $r(t) = 1$ for all $t \leq 16n$ and $r(t) = 2$, otherwise. Figure 5 illustrates an example of the above reduction.

Correctness. First, we provide an intuitive explanation of why this reduction works. We argue that a satisfying assignment to instance ϕ of the 3,4-SAT problem bijects to a satisfying interval selection for the instance $\mathcal{I}_\phi$ of TwIS. The bijection requires that for any literal ℓ that is true in a satisfying assignment of ϕ, the principal interval of ℓ is selected in the solution for $\mathcal{I}_\phi$.

The restrictions imposed by r imply that whenever a principal interval of a literal ℓ is selected, the unit intervals associated with (up to 4) occurrences of $\neg\ell$ cannot be selected (because $r(t) = 1$

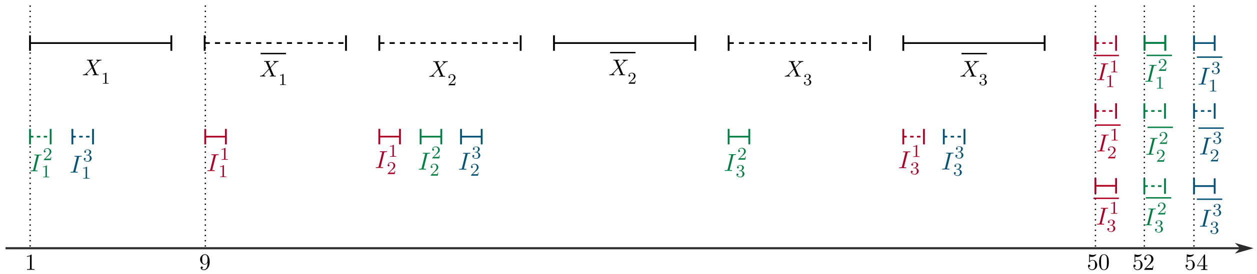


Figure 5: Construction of the instance $\mathcal{I}_\phi$ of the TwIS problem from the 3, 4-SAT instance $\phi = (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3)$. The restriction function $r(t)$ is defined as $r(t) = 1$ for $t \leq 48$ and $r(t) = 2$ otherwise. A satisfying assignment for ϕ is $\sigma(x_1) = 1$ and $\sigma(x_2) = \sigma(x_3) = 0$; the selected intervals in the corresponding solution for $\mathcal{I}$ are shown in solid color.

in their intersecting points). Thus, whenever the principal interval of a literal ℓ is selected (i.e., its twin, which is the principal interval of $\neg\ell$ is not selected), one can select the unit intervals associated with any occurrence of ℓ in clauses where it appears; this is because they intersect with the principal interval of $\neg\ell$, which is not selected. Selecting these unit intervals means their twin unit intervals This yields an equivalent between a satisfying assignment in the SAT formula and the TwIS. We further note that the number of twin intervals in $\mathcal{I}_\phi$ is polynomial in $|X|$ and $|C|$. We can conclude the following lemma, which establishes the hardness of TwIS in the strong sense based on the above reduction.

To establish the main result of this section, we first establish the two sides of the reduction in the following lemmas.

Lemma 4.3. *If the answer to the instance $\mathcal{I}_\phi$ of* TwIS *is yes, then the instance ϕ of* 3, 4-SAT *is satisfiable.*

Proof. Let S be a selection of intervals for $\mathcal{I}_\phi$ that certifies a yes answer to $\mathcal{I}_\phi$. We construct an assignment $\sigma : X \to \{0, 1\}$ for ϕ as follows: For each variable $x_i \in X$, set $\sigma(x_i) = 1$ if the principal interval X_i is in S, and set $\sigma(x_i) = 0$, otherwise. We will prove that σ satisfies ϕ. By construction, for each clause C_j we know that all three unit intervals $\overline{I_k^j}$ (for $k \in [3]$) contain point $t = 16n + 2j$, where $r(t) = 2$. Thus, at least one of these three unit intervals is not included in S. Let $\overline{I_q^j}$ for some $q \in [3]$ be such interval that is not selected; that is, I_q^j is in S. Because I_q^j overlaps with the principal interval of $\neg\ell_q^j$ at some point $t \leq 16n$, and $r(t) = 1$, the principal interval of $\neg\ell_q^j$ cannot be in S. Namely, S contains the principal interval of ℓ_q^j. Therefore, in the corresponding assignment of ϕ, we have $\sigma(\ell_q^j) = 1$, and thus the clause C_j is satisfied. We conclude that each clause C_j in ϕ has at least one true literal with true value σ, and thus σ satisfies ϕ. $\square$

Lemma 4.4. *If an instance $\phi = (X, C)$ of the* 3, 4-SAT *problem is satisfiable, then the answer to the instance $\mathcal{I}_\phi$ of* TwIS *is yes.*

Proof. Let $\sigma : X \to \{0, 1\}$ be a satisfying assignment for ϕ. We construct a solution S for $\mathcal{I}_\phi$, which contains the selected interval from each twin pair. For each variable $x \in X$, if $\sigma(x) = 1$, we select the principal interval of x along with the left (respectively right) intervals of all the unit intervals associated with positive (respectively negative) literals of x. Similarly, if $\sigma(x) = 0$, we select the principal interval of $\neg x$, along with the left (respectively right) intervals of all the unit intervals associated with negative (respectively positive) literals of x.

Next, we prove that S satisfies all restrictions in $\mathcal{I}_\phi$. First, we consider $t \leq 16n$, and show that the restriction imposed by $r(t) \leq 1$ is satisfied. In this range of t, at most two intervals in $\mathcal{I}_t$ contain t. We need to show whenever S includes the principal interval of a variable x, it does not contain the unit intervals that intersect the principal interval of x. This holds because these unit intervals are associated with $\neg x$, which is not selected by the construction of S. Consequently, the number of selected intervals at the intersection points of the principal interval of x and the unit intervals is at most 1, satisfying the restriction $r(t) = 1$ (for $t \leq 16n$).

Next, we consider $t > 16n$ and show the restriction imposed by $r(t) = 2$ is satisfied. There is one unit interval for each literal in a clause C_j, i.e., for any $t > 16n$, there are up to 3 intersecting intervals. Since σ satisfies ϕ, at least one literal in each clause receives a true value. Suppose that it is the q'th literal for $q \in [3]$. As a result, based on the construction of S, $\overline{I_q^j}$ is not selected. This ensures that the number of selected intervals at these points is at most two, satisfying the restriction $r(t) = 2$.

We conclude that the restrictions imposed by $r(t)$ are satisfied for all $t \in [m]$ and thus S is a valid solution for $\mathcal{I}_\phi$. ☐

Lemma 4.5. *Answering instances (I, r, m) of the* TwIS *problem is NP-hard even if the m is polynomial in $|I|$.*

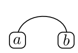
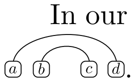
Proof. We will use the reduction from 3,4-SAT that was described in Section 4.1. This reduction takes polynomial time. In particular, given an instance of 3,4-SAT, ϕ with $|X|$ variables and $|C|$ clauses, the number of twin pair intervals in the TwIS is $|X| + 3|C|$, and the domain of r (the horizon) is $[m]$, where $m = 16|X| + 2|C| + 1$, which is polynomial in the number of intervals in the TwIS instance. We conclude that the reduction is polynomial in $|X|$ and $|C|$.

Based on Lemma 4.3 and 4.4, we have established that the 3,4-SAT instance ϕ is satisfiable if and only if the constructed TwIS instance $\mathcal{I}_\phi$ admits a valid selection. ☐

4.2 Hardness of Dome Selection with Prefix Restrictions

We present a polynomial-time reduction from TwIS to a new problem that we refer to as Dome Selection with Prefix Restrictions (DoSePr). Before defining the DoSePr problem, we first define the notion of a *dome* as follows.

Definition 4.6. *A* dome *is formed by four positive integers (a, b, c, d) such that $a \leq b < c \leq d$ and $b - a = d - c$. We refer to (a, d) (respectively (b, c)) as an* arc *with endpoints a and d (respectively b and c). We refer to (a, d) and (b, c) the* outer arc *and* inner arc *of the dome, respectively. When $a = b$ and $c = d$, we call the dome a* singleton dome *and otherwise call it a* regular dome*.*

In our figures, a singleton dome (a, b) is shown as  and regular dome (a, b, c, d) is shown as .

Definition 4.7. *An instance (D, m) of the* Dome Selection with Prefix Restrictions *(*DoSePr*) is defined by a multiset $D = \{D_1, \ldots, D_n\}$ of domes and a positive integer m, where $D_i = (a_i, b_i, c_i, d_i)$ s.t. $d_i \leq m$ for some integer m that we refer to as* horizon*. The decision problem asks whether there is a multiset S of arcs with exactly one arc from each dome D_i, such that, for any $t \in [m]$, it holds that $\mathcal{N}_S(t) \leq t$, where $\mathcal{N}_S(t)$ denotes the number of arc endpoints in S with value at most t, counting each endpoint as many times as it appears across different arcs in S.*

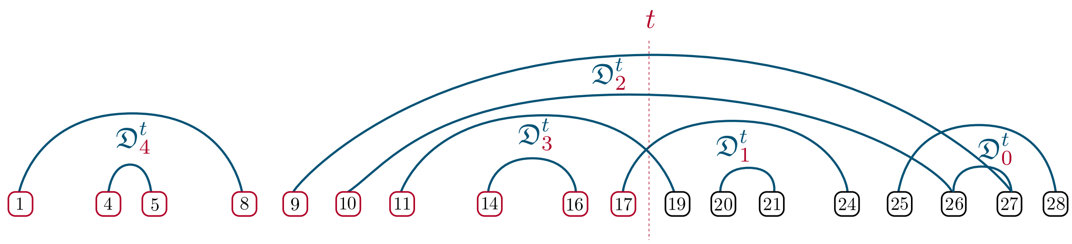


Figure 6: An example of different types of domes based on their positions relative to $t = 18$

Example 1. *Consider an instance of the* DoSePr *with domes* $D = \{(2, 3, 4, 5), (3, 4), (4, 5, 9, 10), (7, 8)\}$ *and* $m = 10$. *A valid solution for this instance is* $S = \{(2, 5), (3, 4), (5, 9), (7, 8)\}$. *For example, when* $t = 5$, *we have* $\mathcal{N}_S(t) = 5$, *which is no more than* t. *In particular, selected endpoints* $\leq t$ *are* $\{2, 5, 3, 4, 5\}$. *Note that there are two endpoints with a value of* 5, *which belong to two different arcs.*

Next, we explain how the TwIS problem reduces to the DoSePr problem.

Construction. Given an arbitrary instance $\mathcal{I} = (I, r, m')$ of TwIS, where $|I| = n'$ for some positive n', we construct the corresponding instance $\mathcal{D} = (D, m)$ of DoSePr as follows. For each twin pair of intervals $(I_i, \overline{I_i}) \in I$, where $I_i = (a', b')$ and $\overline{I_i} = (c', d')$, we construct a regular dome $D_i = (a_i, b_i, c_i, d_i)$ where $a_i = 6n'a', b_i = 6n'b' + 3n', c_i = 6n'c'$, and $d_i = 6n'd' + 3n'$. It is easy to verify that D_i is indeed a dome, that is $b_i - a_i = d_i - c_i$ (see Observation 4.1).

We define m to be a large enough horizon. In particular, we let $m = 164(n'm')^2$. For any $t \in [m]$ that is a multiple of $6n'$ and $t \leq 3n'(2m' + 1)$, we further add $d(t)$ some singleton domes that all start at t and end at m. In other words, we add $d(t)$ identical singleton domes (t, m). Next, we explain how $d(t)$ is defined. For convenience, we let $d(t) = 0$ if t is not a multiple of $6n'$ or $t > 3n'(2m' + 1)$. Suppose t is indeed a multiple of $6n'$. The idea is to define $d(t)$ in a way to project the requirements imposed by the restriction function r in $\mathcal{I}$ to the requirement $\mathcal{N}_S(t) \leq t$ in DoSePr.

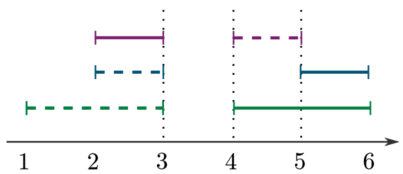
Let $\mathfrak{D}_i^t$ be the set of regular domes with exactly i endpoints before or at point t, and define $c(t) = 2|\mathfrak{D}_4^t| + |\mathfrak{D}_3^t| + |\mathfrak{D}_2^t|$.

Example 2. *For* $t = 18$, *dome* $(1, 4, 5, 8) \in \mathfrak{D}_4^t$, *dome* $(11, 14, 16, 19) \in \mathfrak{D}_3^t$, *dome* $(9, 10, 26, 27) \in \mathfrak{D}_2^t$, *dome* $(17, 20, 21, 24) \in \mathfrak{D}_1^t$, *and dome* $(25, 26, 27, 28) \in \mathfrak{D}_0^t$. *Figure 6 provides an illustration.*

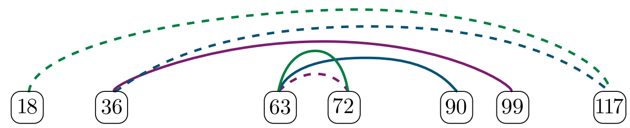
Intuitively, this definition of c implies that there are $c(t)$ arc endpoints with value at most t in any valid solution S (a set of arcs), regardless of the choices made to form S. This is because:

- (i) all arc endpoints of domes in $\mathfrak{D}_4^t$ are at most t, and two of them (associated with one arc) contribute to $c(t)$.
- (ii) three arc endpoints of domes in $\mathfrak{D}_3^t$ are at most t, and any such dome contributes at least 1 to $c(t)$.
- (iii) the left endpoints of both arcs of domes in $\mathfrak{D}_2^t$, and since one arc is in S, the dome contributes exactly 1 to $c(t)$.

Finally, we let



(a) The instance $\mathcal{I}$ of TwIS



(b) The corresponding instance $\mathcal{D}$ of DoSePr (only regular domes are depicted.)

Figure 7: An illustration of the construction of an instance $\mathcal{D}$ of DoSePr from an instance $\mathcal{I}$ of TwIS, as explained in Example 3. Solid lines show the selected intervals in a valid solution of $\mathcal{I}$ and the selected arcs in the corresponding solution in $\mathcal{D}$.

$$d(t) = t - c(t) - \sum_{j<t} d(j) - r(t/(6n')).$$

The scaling argument that is applied when forming the DoSePr instance from TwIS instance ensures that $d(t)$ is indeed non-negative for any $t \in [m]$ (see Observatio 4.2 for details). This completes our construction of the DoSePr instance. In a nutshell, we have added one regular dome per twin interval in the TwIS instance, and for any $t \in [m]$, we added extra identical singleton domes to capture the requirements imposed by the restriction function in the TwIS instance.

Example 3. *Consider an instance* $\mathcal{I} = (I, r, m')$ *of the* TwIS *problem with* $I = \{((1,3)(4,6)), ((2,3),(5,6)), ((2,3),(4,5))\}$ *and* $m' = 6$. *Suppose* $r(1) = r(2) = r(6) = 3$, $r(3) = r(4) = 1$, *and* $r(5) = 2$. *The corresponding instance* $\mathcal{D} = (D, m)$ *of* DoSePr *has regular domes* $D = \{(18,63,72,117),(36,63,90,117),(36,63,72,99)\}$ *and* $m = 164(n'm')^2 = 164 \cdot (3 \cdot 6)^2 = 53{,}136$ *(see Figure 7). In addition,* $3n'(2m'+1) = 117$ *for any* $t \in [1, 117]$ *that is a multiple of* $6n' = 18$, $d(t)$ *singleton domes* (t, m) *are added. Some examples of* $d(t)$ *are shown as follows.*

$$\begin{aligned}
d(18) &= 18 - 0 - 0 - 3 = 15 & c(18) &= 0\\
d(36) &= 36 - 0 - 15 - 3 = 18 & c(36) &= 0\\
d(54) &= 54 - 0 - (18 + 15) - 1 = 20 & c(54) &= 0\\
d(72) &= 72 - 3 - (15 + 18 + 20) - 1 = 15 & c(72) &= 3
\end{aligned}$$

Correctness. First, we provide some intuitions about the correctness of the reduction. We show that any valid solution S for instance $\mathcal{D}$ of DoSePr bijects to a valid solution S' of $\mathcal{I}$. Note that S contains exactly one arc, the inner or the outer arc, of each regular dome D_i of $\mathcal{D}$ (in addition to singleton domes, which are contained in any valid solution for $\mathcal{D}$). Now, S' selects the left interval I_i when the outer arc of D_i is selected in S and the right interval $\overline{I_i}$ when the inner arc of D_i is selected in S (see Figure 7). We let $\Gamma_{S'}(t')$ denote the number of intervals selected by S' that include $t' \in [m']$. We show that S is a valid solution for $\mathcal{D}$ if and only if S' is a valid solution to $\mathcal{I}$.

Suppose S is a valid solution for DoSePr. Then we must have $\mathcal{N}_S(t) \leq t$ for any $t \in [m]$, where $\mathcal{N}_S(t)$ is the number of arc endpoints in S with value at most t. As mentioned earlier, $c(t)$ endpoints will be present in any valid solution regardless of the choices to form S. Moreover, S must contain all endpoints of singleton domes in $\mathcal{D}$ for $j \leq t$. That is, any valid solution for $\mathcal{D}$, including S, must contain $c(t) + \sum_{j \le t} d(j)$ arc endpoints. In addition to these fixed points, there will be more arcs in S, which are contributed by the domes in $\mathfrak{D}_1^t$ and $\mathfrak{D}_3^t$ as follows. Let D_i be a regular dome in $\mathcal{D}$, and let $(I_i, \overline{I}_i)$ be its corresponding twin intervals in $\mathcal{I}$.

- Suppose $D_i \in \mathfrak{D}_1^t$. Now, if S contains the inner arc of D_i, the contribution of D_i to $\mathcal{N}_S(t)$ would be 0. This is equivalent to including the right interval $\overline{I_i}$ in S', and thus the twin interval $(I_i, \overline{I_i})$ does not contribute $\Gamma_{S'}(t')$ for $t' = t/6n$. Informally, the contribution of the dome D_i to the left-hand side of inequality $N_S(t) \le t$ and the contribution of $(I_i, \overline{I_i})$ to the left-hand side of the inequality $\Gamma_{S'}(t') \le r(t')$ will be both 0. Similarly, if S contains the outer arc of D_i, the contribution of D_i to $\mathcal{N}_S(t)$ would be 1. This is equivalent to including the left interval $\overline{I_i}$ in the TwIS instance, and $(I_i, \overline{I_i})$ contribute 1 item to $\Gamma_{S'}(t')$. Informally, the contribution of the dome D_i to the left-hand side of inequality $\mathcal{N}_S(t) \le t$ and the contribution of $(I_i, \overline{I_i})$ to the left-hand side of $\Gamma_{S'}(t/(6n')) \le r(t')$ will be both 1.

- Suppose $D_i \in \mathfrak{D}_3^t$. Assume S contains the inner (respectively the outer) arc of D_i. In this case, in addition to the fixed 1 unit of the contribution of D_i to $\mathcal{N}_S(t)$ (captured in $c(t)$), it further contributes 1 (respectively 0) unit to $\mathcal{N}_S(t)$. This is equivalent to including the right interval $\overline{I}_i$ (respectively the left interval I_i) in S'. In this case, the number of selected intervals in S' that intersect t is increased by 1 (respectively 0). Intuitively, both left-hand sides of $\mathcal{N}_S(t) \le t$ and $\Gamma_{S'}(t/(6n')) \le r(t')$ increase by 1 (respectively 0).

We note that $n = |D|$ is polynomial in both n' and m', and thus $m = 164(n'm')^2$ is polynomial in n. Therefore, we can conclude the following lemma, which establishes the NP-hardness of DoSePr. We start the formal proof with the following observation showing that the regular domes in $\mathcal{D}$ are indeed valid.

Observation 4.1. *For any $i \in [n']$, any constructed $D_i = (a_i, b_i, c_i, d_i)$ in the reduction is a valid dome, i.e., $b_i - a_i = d_i - c_i$.*

Proof. For each twin pair of intervals $I_i = (a_i', b_i')$ and $\overline{I_i} = (c_i', d_i')$ we know that $b_i' - a_i' = d_i' - c_i'$ by the definition of the TwIS problem. Multiplying by $6n'$ and adding $3n'$ to each side, we have $(6n'b_i' + 3n') - 6n'a_i' = (6n'd_i' + 3n') - 6n'c_i'$. Based on the construction of D_i, $(6n'b_i' + 3n') - 6n'a_i' = b_i - a_i$ and $(6n'd_i' + 3n') - 6n'c_i' = d_i - c_i$. ☐

Next, we observe that the number $d(t)$ of extra singleton domes added for any $t \in [m]$ is non-negative.

Observation 4.2. *In the construction of $\mathcal{D}$ from $\mathcal{I}$, for any $t \in [m]$, we have $d(t) \ge 0$.*

Proof. For any t that is not divisible by $6n'$ or $t > 3n'(2m' + 1)$, we have $d(t) = 0$, and the lemma holds for these values of t. For other values of t, which are all $6n'$ or larger, we first establish that $d(t) \ge 5n' - c(t)$. For $t = 6n'$, we have $d(t) = 6n' - c(6n') - r(1) \ge 5n' - c(6n')$ (for any value of t, we have $r(t) \le n'$). In what follows, we assume $t \ge 12n'$ and deduce

$$
\begin{aligned}
d(t) &= t - c(t) - \Sigma_{j<t} d(j) - r(t/(6n')) \\
&= t - c(t) - \Sigma_{j \le t-6n'} d(j) - r(t/(6n')) && (2) \\
&= t - c(t) - d(t-6n') - \Sigma_{j<t-6n'} d(j) - r(t/(6n')) \\
&= t - c(t) - ((t-6n') - c(t-6n') - \Sigma_{j<t-6n'} d(j) - r(\frac{t-6n'}{6n'})) - \Sigma_{j<t-6n'} d(j) - r(t/(6n')) && (3) \\
&= -c(t) + 6n' + c(t-6n') + r(\frac{t-6n'}{6n'}) - r(t/(6n')).
\end{aligned}
$$

In Equality (2), we used the fact that singleton domes are added only for multiples of $6n'$. In Equality (3), we used the definition of $d(t-6n')$. As $c(t-6n')$ and $r(\frac{t-6n'}{6n'})$ are both non-negative, and by definition, $r(t/(6n'))$ is at most n, we can conclude that

$$d(t) \geq -c(t) + 6n' - r(t/(6n')) \geq 5n' - c(t).$$

Moreover, we can bound $c(t)$ as follows:

$$c(t) = 2|\mathfrak{D}_4^t| + |\mathfrak{D}_3^t| + |\mathfrak{D}_2^t| \leq 2(|\mathfrak{D}_4^t| + |\mathfrak{D}_3^t| + |\mathfrak{D}_2^t|) \leq 2n'.$$

Thus, $d(t) \geq 5n' - c(t) \geq 3n' \geq 0$. ☐

For each dome $D_i \in D$, let the boolean variable x_i indicate whether the outer arc of D_i is selected in S or not. Specifically, $x_i = 1$ means the outer arc is selected, while $\neg x_i = 1$ means the inner arc is selected.

The following two lemma express the values of $\mathcal{N}_S(t)$ and $\Gamma_{S'}(t)$ in a way that allows us to relate a solution S for the DoSePr instance $\mathcal{D}$ to its associated solution S' in the TwIS instance $\mathcal{I}$. These lemmas will later help us in proving the correctness of the reduction.

Lemma 4.8. *Let $\mathcal{D} = (D, m)$ be the* DoSePr *problem constructed from an instance $\mathcal{I}$ of the* TwIS *problem, and let S be a multiset of arcs that contains exactly one arc from each dome $D_i \in D$. For any $t \in [m-1]$, the number of arc endpoints in S with value at most t is exactly $\mathcal{N}_S(t) = \sum_{i\in\mathfrak{D}_1^t} x_i + \sum_{i\in\mathfrak{D}_3^t} \neg x_i + |\mathfrak{D}_2^t| + |\mathfrak{D}_3^t| + 2|\mathfrak{D}_4^t| + \sum_{j\leq t} d(j)$.*

Proof. Fix a value of $t \in [m-1]$. We consider five cases, depending on the positions of domes relative to t, to compute the contribution of different domes to $\mathcal{N}_s(t)$. See Figure 6 for an illustration.

- Domes with all the endpoints after t, namely domes that are in $\mathfrak{D}_0^t$, do not contribute to $\mathcal{N}_S(t)$.
- Each dome with exactly 1 endpoint (the left endpoint of the outer arc) before or on t, namely, a dome in $\mathfrak{D}_1^t$, adds exactly one unit to $\mathcal{N}_S(t)$, which happens when the outer arc of D is selected. Thus, the total contribution of domes in $\mathfrak{D}_1^t$ to $\mathcal{N}_S(t)$ is $\sum_{i\in\mathfrak{D}_1^t} x_i$.
- For each dome with exactly 2 endpoints (the left endpoint of each arc) before or on t, namely, domes in $\mathfrak{D}_2^t$, selecting either of them adds one endpoint to $\mathcal{N}_S(t)$. Therefore, the contributions of domes in $\mathfrak{D}_2^t$ to $\mathcal{N}_S(t)$ is exactly $|\mathfrak{D}_2^t|$.
- Consider a dome D with exactly 3 endpoints before or on t, namely, a dome in $\mathfrak{D}_3^t$. The left endpoint of the arc of D that is present in S contributes 1 unit to $\mathcal{N}_S(t)$ (regardless of the outer or inner arc being present in S). Moreover, if the inner arc of D is selected, there will be another unit of contribution. Therefore, the contributions of domes in $\mathfrak{D}_3^t$ to $\mathcal{N}_S(t)$ is exactly $|\mathfrak{D}_3^t| + \sum_{i\in\mathfrak{D}_3^t} \neg x_i$.
- For each dome with all 4 endpoints before or on t, namely domes in $\mathfrak{D}_4^t$, for each of the arcs, its two endpoints are before or on t, thus domes in $\mathfrak{D}_4^t$ contribute $2|\mathfrak{D}_4^t|$ endpoints to $\mathcal{N}_S(t)$.

Adding the singleton domes added before or at point t, which is $d(t)$, we can write $\mathcal{N}_S(t) = \sum_{i\in\mathfrak{D}_1^t} x_i + \sum_{i\in\mathfrak{D}_3^t} \neg x_i + |\mathfrak{D}_2^t| + |\mathfrak{D}_3^t| + 2|\mathfrak{D}_4^t| + \sum_{j\leq t} d(j)$. ☐

Lemma 4.9. *Let $\mathcal{D} = (D, m)$ be the* DoSePr *problem constructed from an instance $\mathcal{I} = (I, r, m')$ of the* TwIS *problem, and let S' be a selection of intervals from $\mathcal{I}$ that contains exactly one interval from each twin interval $(I_i, \overline{I_i}) \in I$. For any $t' \in [m-1]$, we have $\Gamma_{S'}(t') = \sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i$, where $t = 6n't'$.*

Proof. Consider a regular dome $D_i = (a, b, c, d)$ added in $\mathcal{D}$ for the twin intervals (a', b') and (c', d') in $\mathcal{I}$. For a given $t' \in [m-1]$, let $t = 6t'n$. Recall that $c = 6n'c'$, and $d = 6n'd' + 3n$ by the construction of $\mathcal{D}$ from $\mathcal{I}$. One can assert that $t' \in [a', b']$ if and only if $D_i \in \mathfrak{D}_1^t$. Therefore, the term $\sum_{i \in \mathfrak{D}_1^t} x_i$ corresponds to the twin intervals in $\mathcal{I}$ whose left interval contributes to $\Gamma_{S'}(t')$. Similarly, one can assert that $t' \in [c', d']$ if and only if $D_i \in \mathfrak{D}_3^t$. Therefore, the term $\sum_{i \in \mathfrak{D}_3^t} \neg x_i$ represents the domes in $\mathfrak{D}_3^t$ for which the inner arc is selected; this corresponds to the twin intervals in $\mathcal{I}$ whose right interval contributes to $\Gamma_{S'}(t')$. Summing the two terms, we get $\Gamma_{S'}(t') = \sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i$. ☐

Now, we are ready to prove the correctness of the reduction.

Lemma 4.10. *If the answer to the instance $\mathcal{D} = (D, m)$ of* DoSePr *is yes, then the answer to its corresponding instance $\mathcal{I} = (I, r, m')$ of* TwIS *is also yes.*

Proof. Consider a solution S that certifies a yes answer for the instance $\mathcal{D}$ of DoSePr. We will construct a solution S' for $\mathcal{I}$ from S and prove that it is a valid solution for $\mathcal{I}$. For each regular dome D_i in $\mathcal{D}$ and its corresponding twin intervals $(I_i, \overline{I_i})$ in $\mathcal{I}$, if the inner arc of D_i is selected in S, we select $\overline{I_i}$ in S'. Similarly, if the outer arc of D_i is selected by S, we select I_i in S'.

We need to show that for each $t' \in [m']$, the number of selected intervals in S' intersecting t' is at most $r(t')$. Let $t = 6n't'$. Since S is a valid solution for $\mathcal{D}$, it holds that $\mathcal{N}_S(t) \leq t$. Using Lemmas 4.8 and 4.9, we can write the following.

$$
\begin{aligned}
\Gamma_{S'}(t') &= \sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i && \text{Lemma 4.9} && (4)\\
&= \mathcal{N}_S(t) - (|\mathfrak{D}_2^t| + |\mathfrak{D}_3^t| + 2|\mathfrak{D}_4^t| + \sum_{j \leq t} d(j)) && \text{Lemma 4.8} && (5)\\
&\leq t - (|\mathfrak{D}_2^t| + |\mathfrak{D}_3^t| + 2|\mathfrak{D}_4^t| + \sum_{j \leq t} d(j)) && \mathcal{N}_S(t) \leq t && (6)\\
&\leq t - c(t) - \sum_{j \leq t} d(j) && \text{by definition of } c(t) && (7)\\
&= t - c(t) - \sum_{j < t} d(j) - d(t) && &&\\
&= t - c(t) - \sum_{j < t} d(j) - (t - c(t) - \sum_{j < t} d(j) - r(t/(6n'))) && \text{by definition of } d(t) &&\\
&= r(t'). && && (8)
\end{aligned}
$$

Therefore, $\Gamma_{S'}(t') \leq r(t')$ for any $t' \in [m]$ and thus S' is a valid solution for instance $\mathcal{I}$. ☐

Lemma 4.11. *If the answer to the instance $\mathcal{I} = (I, r, m')$ of* TwIS *is yes, then the answer to its corresponding instance $\mathcal{D} = (D, m)$ of* DoSePr *is also yes.*

Proof. Consider a solution S' that certifies a yes answer for the instance $\mathcal{I}$ of TwIS. We will construct a solution S from S' and prove that it is a valid solution for $\mathcal{D}$. For each regular dome D_i in $\mathcal{D}$ and its corresponding twin interval pair $(I_i, \overline{I_i})$ in $\mathcal{I}$, if the left interval I_i is included in S', we include the outer arc in S, and if the right interval $\overline{I_i}$ is in S' we include the inner arc in S.

Clearly, all singleton intervals are also included in S. This ensures that exactly one arc from each dome is in S.

In order to show that S is a valid solution for $\mathcal{D}$, we will show that for each $t \in [m]$, $\mathcal{N}_S(t) \leq t$. By the construction of $\mathcal{D}$, all singleton domes find their right endpoints at m. Let $q = 6n'm' + 3n'$ and observe that all arcs of regular domes find their endpoint at $\leq q$. We establish $\mathcal{N}_S(t) \leq t$ using different arguments depending on the value of t relative to q.

- Suppose $t \in [1, q]$ and t is a multiple of $6n'$. Since S' is a valid solution for $\mathcal{I}$, it must hold that, for each $t' \in [m']$, $\Gamma_{S'}(t') \leq r(t')$. consider the points that ares Based on Lemma 4.8, we can rewrite the inequality on point $t = 6n't'$ in the DoSePr instance as follows.

$$\begin{aligned}
\mathcal{N}_S(t) &= \sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i + (|\mathfrak{D}_2^t| + |\mathfrak{D}_3^t| + 2|\mathfrak{D}_4^t| + \sum_{j \leq t} d(j)) && \text{Lemma 4.8} && (9)\\
&= \Gamma_{S'}(t') + (c(t) + \sum_{j \leq t} d(j)) && \text{Lemma 4.9} && (10)\\
&\leq r(t') + c(t) + \sum_{j \leq t} d(j) && \Gamma_{S'}(t') \leq r(t') && (11)\\
&= r(t') + c(t) + \sum_{j < t} d(j) + d(t) && &&\\
&= r(t') + c(t) + \sum_{j < t} d(j) + (t - c(t) - \sum_{j < t} d(j) - r(t/(6n'))) && \text{by definition of } d(t) &&\\
&= t. && && (12)
\end{aligned}$$

- Suppose $t \in [1, q]$ and t is multiple of $3n'$ but not a multiple of $6n'$. Let $t_p = t - 3n'$ and note that t_p is a multiple of $6n'$. Therefore, we can write (from the previous case) $\mathcal{N}_S(t_p) \leq t_p$. We claim that $\mathcal{N}_S(t) \leq \mathcal{N}_S(t_p) + 2n'$. By Lemma 4.8, $\mathcal{N}_S(t) = \sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i + c(t) + \sum_{j \leq t} d(j)$. Moreover, we have $\sum_{j \leq t} d(j) = \sum_{j \leq t_p} d(j)$ because singleton domes start only in multiples of $6n'$. On the other hand, the sum of the three other terms, namely, $\sum_{i \in \mathfrak{D}_3^t} \neg x_i + \sum_{i \in \mathfrak{D}_1^t} x_i + c(t)$ is at most $2n'$ for any t (because the number of regular domes is exactly n'). Therefore, $\mathcal{N}_S(t) - \mathcal{N}_S(t_p) \leq 2n'$ (in an extreme case, when the sum of the three terms is 0 in t_s and $2n'$ in t, the difference between $\mathcal{N}_S(t)$ and $\mathcal{N}_S(t_p)$ will be $2n'$). We can conclude that $\mathcal{N}_S(t) \leq \mathcal{N}_S(t_p) + 2n' \leq t_p + 2n' \leq t$.

- Suppose $t \in [1, q]$ and t is not a multiple of $3n'$. Let t_s be the largest multiple of $3n'$ that is smaller than t. Since all arcs endpoints are on multiple of $3n'$, we can write $\mathcal{N}_S(t) = \mathcal{N}_S(t_s) \leq t_s < t$. The first inequality holds because we established $\mathcal{N}_S(t_s) \leq t_s$ for values of t_s that are multiples of $3n'$ in previous cases.

- Suppose $t \in (q, m)$. Based on our construction, there is no arc endpoint at t' where $q < t' < m$. Thus, in order to show $\mathcal{N}_S(t) \leq t$, it suffices to show that $\mathcal{N}_S(q) \leq q$, which we will establish in the next case (where $t \leq q$)

- Suppose $t = m$. The contributions of each dome to $\mathcal{N}_S(t)$ is exactly 2. Given that we have n' regular domes and $\sum_{t \in [q]} d(t)$ singleton domes, we can write $\mathcal{N}_S(m) = 2n' + \sum_{t \in [q]} 2d(t)$. Therefore, the following holds.

$$
\begin{aligned}
\mathcal{N}_S(m) &= 2n' + \sum_{t\in[q]} 2d(t) \\
&\leq 2n' + \sum_{t\in[q]} 2t && d(t) \leq t \\
&\leq 2n' + 2q^2 \\
&= 2n' + 2(6n'm' + 3n')^2 && q = 6n'm' + 3n' \\
&\leq 164(n'm')^2 = m.
\end{aligned}
$$

In conclusion, we have $\mathcal{N}_S(t) \leq t$ for all values of $t \in [m]$, and we can conclude S certifies that $\mathcal{D}$ is a yes instance of the DoSePr problem. ☐

From the above lemmas, we can conclude the main result of this section.

Lemma 4.12. *Answering instances (D, m) of the* DoSePr *problem is NP-hard even if m is polynomial in $|D|$.*

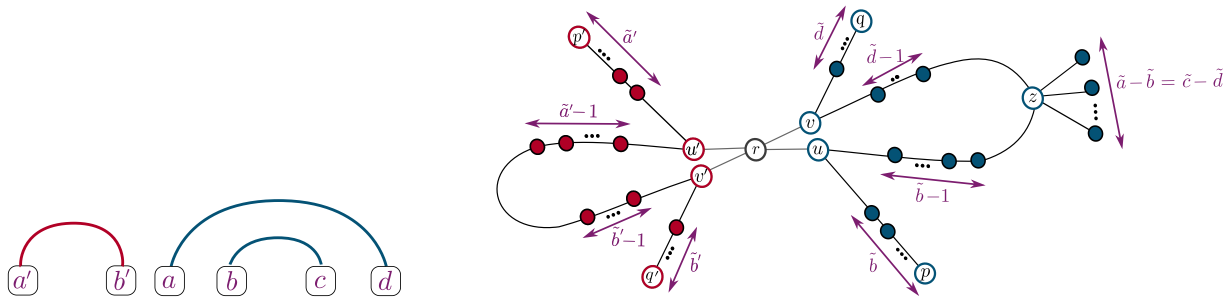
Proof. We note that $n = |D|$ is polynomial in both n' and m'. Moreover, $m = 164(n'm')^2$, which is polynomial in n' and m' (and thus n). The reduction takes polynomial time in m and n and hence is polynomial in m' and n'. Finally, based on Lemma 4.10 and 4.11, we have established that the instance $\mathcal{I} = (I, r, m')$ is a yes-instance of TwIS if and only if the corresponding instance $\mathcal{D} = (D, m)$ is a yes-instance of the DoSePr problem. This completes the proof of equivalence and validates the reduction from TwIS to DoSePr. ☐

4.3 Hardness of Telephone Broadguess in Snowflake Graphs

We refer to the decision variant of the broadcasting problem as the Telephone Broadguess problem with instances (G, s, ρ), where G is the input graph, s is the source, and ρ is a positive integer. The decision question asks whether it is possible to complete broadcasting in G from s within ρ rounds. This section presents our final reduction, from DoSePr to Telephone Broadguess in snowflake graphs.

Construction. Given an instance $\mathcal{D} = (D, m)$ of DoSePr we construct an instance $\mathcal{T} = (G, r, \rho = 2m)$ of Telephone Broadguess, where G is a snowflake graph with center r. For a given $x \in [\rho]$, let $\tilde{x} = \rho - x$. We construct the graph G by creating a reduced caterpillar (see Definition 2.3) for each dome in $\mathcal{D}$ as follows.

- The reduced caterpillar of a singleton dome (a, b) is a path of length $2\tilde{a} + 2\tilde{b}$ with endpoints p and q (here, p, q, z are special vertices in the reduced caterpillar, where z is any arbitrary vertex).
- The reduced caterpillar of a regular dome (a, b, c, d) is formed by a path of length $2\tilde{b} + 2\tilde{d} + 1$ with endpoints p and q, which are two special vertices of the reduced caterpillar. The other special vertex z of the reduced caterpillar is the vertex at distance $2\tilde{b}$ of p (and thus distance $2\tilde{d}$ of q). In addition to the vertices on the path between p and q, the reduced caterpillar includes $\tilde{a} - \tilde{b}$ (which equals $\tilde{c} - \tilde{d}$) other vertices which find z as their sole neighbor.



(a) An instance $\mathcal{D}$ of the DoSePr instance

(b) The corresponding instance $\mathcal{T} = (G, r, \rho)$ of the Telephone Broadguess constructed from $\mathcal{D}$.

Figure 8: An illustration of the construction of the Telephone Broadcasting instance corresponding to a DoSePr instance

For any reduced caterpillar C constructed above (from either a singleton or a regular dome), we label the vertex at distances $\tilde{a}$ form p as the *first gate* of C, denoted as u, and the one at distance $\tilde{b}$ from q as the *second gate of* C and denote it with v. We form the equivalent instance of the Telephone Broadguess problem by forming a graph G with a center vertex r (source) that is connected to all gates of all reduced caterpillars formed by the domes of D. By Definition 2.3, G will be a snowflake graph with center r. Figure 8 provides an illustration of this reduction.

Correctness. Before proving the correctness of our reduction, we prove the following lemmas for broadcasting in the reduced caterpillars associated with singleton and regular domes, respectively.

Lemma 4.13. *Consider a reduced caterpillar C of a singleton dome $D = (a, b)$ with gates u and v. One can complete broadcasting in C within ρ rounds if and only if u is informed no later than time a and v is informed no later than time b.*

Proof. First, suppose the first gate u is informed at time $a' \leq a$ and v is informed at time $b' \leq b$. We explain how broadcasting in C can be completed within ρ rounds. For that, u (respectively v) first informs its neighbor on its path to the special vertex p (respectively q) and then its other neighbor on its path towards v (respectively u). Given that the distance of u to p is $\tilde{a}$ (respectively the distance of v to q is $\tilde{b}$), all vertices on the path between u and p (respectively v to q) will be informed by round $a + \tilde{a} = \rho$ (respectively $b + \tilde{b} = \rho$). See Figure 9 for an illustration.

Next, assume broadcasting in C has been completed by ρ rounds. Given that p (respectively q) must receive the message through u (respectively v), it must be that u (respectively v) is informed by round $\rho - \tilde{a} = a$ (respectively $\rho - \tilde{b} = b$). □

Lemma 4.14. *Consider a reduced caterpillar C of a regular dome $D = (a, b, c, d)$ with gates u and v. One can complete broadcasting in C within ρ rounds if and only if one of the following happens:*

- *(i) u is informed by time b and v is informed by time c.*
- *(ii) u is informed by time a and v is informed by time d.*

Proof. First, we show that if either (i) or (ii) hold, then broadcasting completes within ρ rounds.

- Suppose (i) holds, that is, u and v are informed by times b and c, respectively. We describe a broadcast scheme S that completes within ρ rounds. See Figure 10a for an illustration.

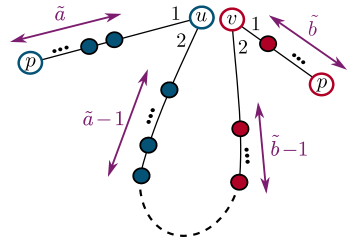


Figure 9: The broadcast scheme of the reduced caterpillar associated with a singleton dome (a, b). Vertex u (respectively v) is responsible for informing the vertices highlighted in blue (respectively red).

In S, vertex u first informs the neighbor on its path towards p, and then it informs the neighbors on its path towards z. This means that, by round ρ, all $\tilde{b}$ vertices on the path between u and p will be informed. Similarly, $\tilde{b}-1$ vertices on the path between u and z are informed (this excludes z). On the other hand, v first informs the neighbor on its path towards z and then the neighbor on its path towards q. This means that, by round ρ, all $\tilde{d}$ vertices on the path between v and q will be informed; this is because the number of rounds between the time that v is informed and ρ is $\tilde{c}$, and since v first informs the neighbor towards z, $\tilde{c}-1$ rounds remain for informing vertices between v and q. This number of rounds is sufficient because the length of the path between v and q is $\tilde{d} \leq \tilde{c}-1$. This way, z will receive the message at time $c+\tilde{d}$, and assuming z informs its uninformed leaves in arbitrary order, the broadcasting in S completes within $c+\tilde{d}+\tilde{c}-\tilde{d}=\rho$ rounds.

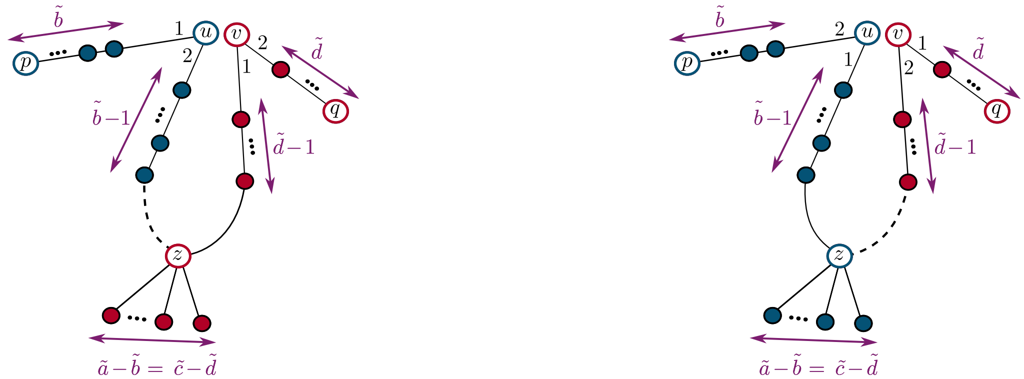
- Next, assume (ii) holds, that is, u and v are respectively informed by rounds a and d. See Figure 10b for an illustration. In this case, u first informs its neighbor on the path towards z and then its other neighbor on the path towards p. This means that by round $a+\tilde{b}+1 \leq a+\tilde{a}=\rho$, all vertices on the path between a and p will be informed. Similarly, by round a, all the $\tilde{b}$ vertices on the path between u and z (including z) will be informed. Thus, all neighbors of z will be informed by $a+\bar{b}+\bar{a}-\bar{b}=\rho$.

  On the other hand, v first informs its neighbor on the path toward q and then its neighbor on the path towards z (excluding z). Thus, all vertices on the path between v and q will be informed by $d+\tilde{d}=\rho$ while vertices on the path between v and z (excluding z) will be informed by time $d+1-(\tilde{d}-1)=\rho$.

Next, assume that there is a broadcast scheme S that completes within ρ rounds. We want to show either (i) or (ii) holds. First, note that in the broadcast tree of S, vertices u and v cannot be in an ancestor-descendent relationship. Otherwise, if u is an ancestor of v (respectively v is an ancestor of u), then q (respectively p) receives the message no earlier than round $\tilde{b}+2\tilde{d}>\rho$ (respectively $2\tilde{b}+\tilde{d}>\rho$). Second, we note that u (respectively v) cannot receive the message later than round b (respectively d); otherwise p (respectively q) will receive the message after round ρ. Moreover, since z has $\tilde{a}-\tilde{b}$ (which equals $\tilde{c}-\tilde{d}$) neighbor that all are leaves, it must receive the message by round $\rho-(\tilde{a}-\tilde{b})=a+\tilde{b}$ (which equals $\rho-(\tilde{c}-\tilde{d})=c+\tilde{d}$) to complete broadcasting within ρ rounds. This means that either u is informed by round a or v is informed by round c (otherwise, z will be informed too late). We conclude that either (i) u is informed by a and v is informed no later than d or (ii) u is informed by b and v is informed by c. ☐

We are now ready to prove the main result of this section.

Theorem 4.15. Telephone Broadcasting *problem is NP-complete for snowflake graphs.*



(a) u and v are informed at rounds b and c, respectively. This corresponds to selecting arc (b, c) in the associated dome.

(b) u and v are informed at rounds a and d, respectively. This corresponds to selecting arc (a, d) in the associated dome.

Figure 10: Two possibilities for informing vertices of a reduced caterpillar (associated with a regular dome). The gates are u and v, where vertex u (respectively v) is responsible for informing the vertices highlighted in blue (respectively red).

Proof. We use the construction mentioned in Section 4.3 to reduce any instance $\mathcal{D} = (D, m)$ of DoSePr to an instance $\mathcal{T} = (G, r, \rho)$. Note that each dome D_i in $\mathcal{D}$ is bijected to a reduced caterpillar C_i in G. The two possibilities for selecting an arc from a regular dome in D_i translate to two possibilities for informing the gates of C_i.

Assume there is a valid solution S for $\mathcal{D}$ in DoSePr. We will show there is a broadcasting scheme S' for $\mathcal{T}$ that completes within ρ rounds. For that, we sort all endpoints of all arcs included in S in the non-decreasing order. Consider any dome $D_i \in D$, and let e_1 and e_2 be the endpoints of the arc selected from D_i in S. Assume e_1 and e_2 have ranks i_1 and i_2 in the sorted order. In the broadcast scheme S', the source r informs the gates of reduced caterpillar C_i associated with D_i at rounds i_1 and i_2. Given that ranks of all arc endpoints in S are distinct, r informs at most one neighbor at each given round.

Next, we show that S' completes within round ρ. Given that S is a valid solution for $\mathcal{D}$, for any $t \in m$, we have $\mathcal{N}_S(t) \leq t$, where $\mathcal{N}_S(t)$ is the number of arc endpoints in S that are at most t. On the other hand, for any e that is the endpoint of an arc in S, we have $\text{rank}(e) \leq \mathcal{N}_S(e)$ (be definition, $\text{rank}(e)$ is upper bounded by the number of arc endpoints in S that are at most e, while $\mathcal{N}_S(e)$ is exactly the number of endpoints that are at most e). We conclude that $\text{rank}(e) \leq e$. Therefore, the gates of any reduced caterpillar C_i in G associated with a singleton dome $D_i = (a, b)$ in D receive the message in S' by rounds (a, b). Similarly, the gates of any reduced caterpillar C_i associated with a regular dome $D_i = (a, b, c, d)$ receive the message in S' by rounds (a, d) (if the arc $(a, d) \in S$) or by rounds (b, c) (if the arc (b, c) is in S). Therefore, by Lemma 4.13 and 4.14, broadcasting in S' completes by round ρ.

For the other side of the reduction, suppose there is a broadcast scheme S' for $\mathcal{T}$ that completes within ρ rounds. We will explain how to construct a valid solution S for $\mathcal{D}$. Consider any reduced caterpillar C_i in G associated with a dome $D_i \in D$. Let (α_i, β_i) be the rounds that center r informs the gateways of C_i. Now, if $D_i = (a, b)$ is a singleton dome, by Lemma 4.13, it must be that the gates of C_i are informed by rounds (a, b), that is $\alpha \leq a$ and $\beta \leq b$. In this case, the single arc of D_i is included in S. Next, if $D_i = (a, b, c, d)$ is a regular dome, by Lemma 4.14, it must be that the gates of C_i are informed by either rounds (a, d) or (b, c), that is either ($\alpha \leq a$ and $\beta \leq d$) or ($\alpha \leq b$ and $\beta \leq c$) must hold. We will include arc (a, d) in S in the former case and (b, c) in the latter case.

In other words, any arc endpoint x in S is associated with a gate that is informed within round x in S'. To show that S is a valid solution for $\mathcal{D}$, we show that for any $t \in [m]$, we have $\mathcal{N}_S(t) \leq t$. Consider the set of endpoints in D that contribute to $\mathcal{N}_S(t)$. As mentioned above, any of these endpoints is associated with a gate that is informed by r within round t. Given that r informs up to t gates by round t, it must be that $\mathcal{N}_S(t) \leq t$. We conclude that S is a valid instance of $\mathcal{D}$.

One can verify that the size of the graph G in the TELEPHONE BROADGUESS instance is polynomial on $n = |D|$ and m. Therefore, given that the DOSEPR is NP-hard by Lemma 4.12, we conclude that TELEPHONE BROADGUESS is NP-hard. TELEPHONE BROADCASTING in general graphs is known to be in NP [24]. Together, these results establish the NP-completeness of TELEPHONE BROADGUESS in snowflake graphs. ☐

5 Constant-Factor Approximation for Bounded Pathwidth

Graphs of pathwidth 1 are caterpillars [21], which are special types of trees, for which the TELEPHONE BROADCASTING problem is solvable in linear time [8]. On the other hand, for graphs with a pathwidth larger than 1, the TELEPHONE BROADCASTING problem is NP-hard, as established by our result for the hardness of the problem in snowflake graphs (Theorem 4.15). Recall that snowflake graphs have pathwidth 2 (Observation 2.1).

In this section, we establish the existence of a constant-factor approximation for TELEPHONE BROADCASTING on graphs with bounded pathwidth. Recall that we use $\boldsymbol{br}^*(G, s)$ to denote the optimal broadcasting time for an instance (G, s). We will demonstrate that the algorithm of Elkin and Kortsarz [7], which has an approximation factor of $\mathcal{O}\left(\frac{\log n}{\log \boldsymbol{br}^*(G,s)}\right)$ for any graph G, achieves a constant factor approximation for graphs of bounded pathwidth. For general graphs, this algorithm has an approximation factor of $\mathcal{O}(\log n/\log\log n)$ (because $\boldsymbol{br}^*(G, s) \geq \log n$), which is the best known approximation factor.

For any graph G of pathwidth w, we will show that $\boldsymbol{br}^*(G, s) = \Omega(n^{4^{-(w+1)}})$, which establishes that the algorithm of Elkin and Kortsarz [7] has a constant factor approximation for graphs of bounded pathwidth w.

To find a lower bound for $\boldsymbol{br}^*(G, s)$ where G is a graph of constant pathwidth, we repeatedly remove a vertex from G and show that the broadcasting in the remainder of G is not much slower compared to G. For that, we will use the following lemma, which holds for any instance of TELEPHONE BROADCASTING (but we only use it for graphs of bounded pathwidth).

Lemma 5.1. *Consider any instance (G, s) of* TELEPHONE BROADCASTING*, and let v be any arbitrary vertex in G. Let $H_1, \ldots, H_m$ be the connected components resulting from removing v from G $(m \geq 1)$. For any $i \in [m]$, choose an arbitrary vertex $s_i \in H_i$. Then, the following inequality holds: $\sum_{i\in[m]} \boldsymbol{br}^*(H_i, s_i) \leq \boldsymbol{br}^*(G, s)(2\boldsymbol{br}^*(G, s) + 1)$.*

Proof. Consider the optimal broadcast tree T^* that completes broadcasting in (G, s) in $\boldsymbol{br}^*(G, s)$ rounds. Suppose v has d children $u_1, \ldots, u_d$ in T^* for some $d \geq 0$. Removing v from T^* results in $d + 1$ subtrees, $\tau_0, \ldots, \tau_d$, where τ_0 is rooted at s and τ_i is rooted at u_i for each $i \in [1, d]$. Note that all vertices in any fixed τ_j belong to the same connected component H_i. Let $\tau_1^i, \ldots, \tau_{p_i}^i$ be the subtrees that form the connected component H_i for some $p_i \geq 0$. The indexing is defined in a way to ensure that $s_i \in \tau_1^i$, and there is an *auxiliary edge* between a vertex in τ_j^i and a vertex in $\tau_{j'}^i$ for some $j' \leq j - 1$ (this is possible because H_i is connected). Figure 11 provides an illustration.

A broadcasting scheme for H_i can be formed as follows:

- The message is transmitted from s_i to the root of τ_1^i within $\boldsymbol{br}^*(G, s)$ rounds.

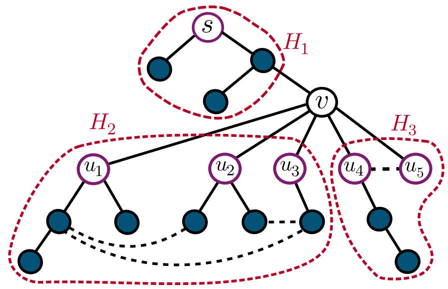


Figure 11: An illustration of the proof of Lemma 5.1. Auxiliary edges are dotted. In this example, the connected components H_1 to H_3 formed after removing v are shown.

- Broadcasting within τ_1^i completes in at most $\boldsymbol{br}^*(G, s)$ rounds, in a similar way that the message is broadcasted in T^*.
- The message is sent to the next subtree τ_2^i along an auxiliary edge in one extra round.

The above process takes at most $2\boldsymbol{br}^*(G, s) + 1$ rounds. Similarly, each subtree τ_j^i takes at most $2\boldsymbol{br}^*(G, s)+1$ rounds to receive the message and broadcast it within τ_j^i. Using this procedure iteratively, broadcasting within (H_i, s_i) completes in at most $p_i(2\boldsymbol{br}^*(G, s) + 1)$ rounds. Summing over all H_i, we get

$$\sum_{i\in[m]} \boldsymbol{br}^*(H_i, s_i) = (2\boldsymbol{br}^*(G, s) + 1) \sum_{i\in[m]} p_i.$$

Note that $\sum_{i\in[m]} p_i$ is a lower bound for $\boldsymbol{br}^*(G, s)$ since it is the degree of v in T^*. Then we conclude

$$\sum_{i\in[m]} \boldsymbol{br}^*(H_i, s_i) \leq \boldsymbol{br}^*(G, s)(2\boldsymbol{br}^*(G, s) + 1).$$

□

For completeness, we show that Lemma 5.1 is asymptotically tight.

Observation 5.1. *There are instances of the* Telephone Broadcasting *problem* (G, s) *and vertex* $v \in G$ *for which Lemma 5.1 is asymptotically tight.*

Proof. We provide (G, s) and $v \in G$ such that if removing v from G results in connected components $H_1, \ldots, H_m$, then $\sum_{i\in[m]} \boldsymbol{br}^*(H_i, s_i) = \Omega(\boldsymbol{br}^*(G, s)^2)$. Let G be a *fan graph* formed by a path P of $n-1$ vertices, where each vertex $i \in [1, n-1]$ is connected to a *center* vertex c. Suppose $v = c$ and s is an endpoint of P. We claim that $\boldsymbol{br}^*(G, s) = \mathcal{O}(\sqrt{n})$. This broadcast time can be achieved by a broadcast tree in which v has $\Theta(\sqrt{n})$ children, and any neighbor of v in the tree is responsible for informing $\Theta(\sqrt{n})$ other vertices (see Figure 12). On the other hand, $G \setminus \{v\}$ has only one connected component, which is a path H_1 in which, setting the source as u_1 gives $\sum_{i\in[m]} \boldsymbol{br}^*(G_i, s_i) = \boldsymbol{br}^*(H_1, u_1) = n - 1 \in \Omega(\boldsymbol{br}^*(G, s)^2)$. □

The main result of this section can be stated as follows.

Theorem 5.2. *There is a polynomial-time algorithm for* Telephone Broadcasting *in graphs with constant pathwidth* w*, achieving an approximation ratio of* $\mathcal{O}(4^w)$*.*

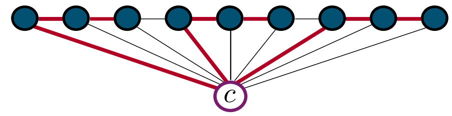


Figure 12: An illustration of Observation 5.1

Proof Overview. As mentioned above, it suffices to prove a lower bound of $\Omega(n^{4^{-(w+1)}})$ for $\boldsymbol{br}^*(G,s)$. For that, we assume G is a "w-path" in the sense that any two vertices that share a bag are neighbors. If G is not a w-path, we will add the missing edges to get a w-path without increasing the pathwidth. Clearly, the addition of new edges does not decrease the broadcast time of G, and since we are looking for a lower bound for the broadcast time, it suffices to focus on w-paths.

We assume a *standard path decomposition* of the input graph G, in which no bag is a subset of another. In such decompositions, we define the notion of *span* of a vertex v as the number of bags where v appears. The span of a decomposition is then the maximum span over the spans of all its vertices.

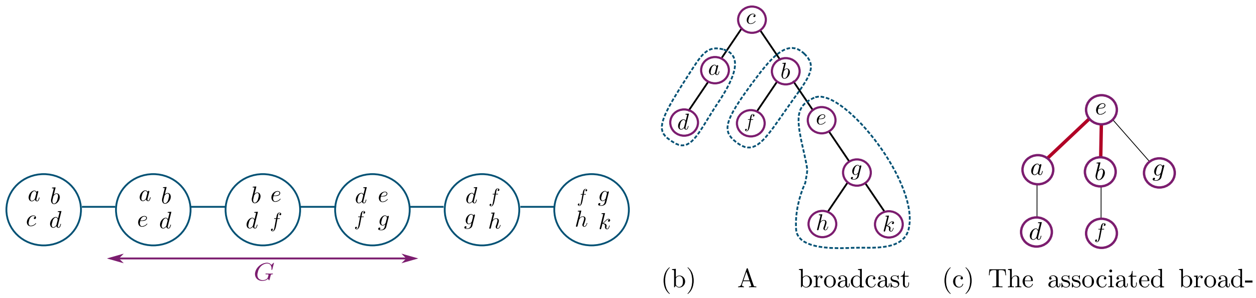
The proof of lemma is established by induction over the size of the input graph G. For that, we consider two cases. First, if the span of the decomposition is "small", we argue that the diameter of G will be "large", and then the desired lower bound for $\boldsymbol{br}^*(G,s)$ holds. On the other hand, when the span is "large", we consider the vertex v_m that has the maximum span and consider the graph G_{v_m} induced by vertices that share a bag with v_m. Note that G_{v_m} is a smaller graph compared to G, and we can show that broadcasting in G_{v_m}, starting from any vertex, cannot be much slower than broadcasting in G, starting from s (Lemma 5.3). Moreover, we will use Lemma 5.1 to show that if we extract v_m from G_{v_m} to get a set $\{H_1,\ldots,H_q\}$ of disjoint connected components, total broadcast time in these component components (starting from arbitrary sources) is not much slower in the absence of v_m. On the other hand, since v_m is removed from G_{v_m}, any of these components H_i has a pathwidth that is at least one unit less than G_{v_m}, and we can use an inductive argument to achieve a lower bound on the broadcast times of any H_i (Lemma 5.4). In summary, by the induction hypothesis, total broadcast time in H_i's gives a lower bound for broadcasting in G_{v_m} which itself gives a lower bound for broadcasting in G.

Now for the formal proof, we prove a lower bound for the broadcast time of any connected graph G with n vertices and pathwidth $w \in \mathcal{O}(1)$. In particular, for any $s \in G$, we will show $\boldsymbol{br}^*(G,s) \in \Omega(n^{4^{-w}})$. Consider a fixed path decomposition of G; we assume the bags in the deposition are arranged from left to right. For now, suppose s is in the left-most bag (this assumption will be relaxed later in Lemma 5.4).

We assume the path decomposition is "standard" in the sense that no bag is a subset of another bag (otherwise, one can remove the smaller bags to attain a standard decomposition). Finally, we assume that G is a "w-path" in the sense that any two vertices located in the same bag are connected in G (all edges allowed in the decomposition are present). If the input graph is not a w-path, we can make it w-path by adding all the missing edges. Clearly, the addition of new edges does not decrease the broadcast time of G, and since we are looking for a lower bound for the broadcast time, it suffices to focus on w-paths.

We now proceed with a formal proof.

Lemma 5.3. *Let (H,s) be an instance of the broadcast problem, where H is a connected w-path. Suppose we have a standard path decomposition D of width w for H where s appears in the first bag of D. Let G be an induced subgraph of H formed by vertices that appear in a consecutive set of bags in D and suppose a vertex $s' \in G$ appears in all such bags. Then, we can write*



(a) Standard path decomposition for a graph H

(b) A broadcast scheme T_H for (H, s)

(c) The associated broadcast scheme T_G for (G, s')

Figure 13: An illustration of Lemma 5.3

$\boldsymbol{br}^*(G, s') \leq \boldsymbol{br}^*(H, s) + 2w$.

Proof. Suppose T_H is an optimal broadcast scheme for the instance (H, s). We will construct a broadcast scheme T_G for (G, s') as follows. See Figure 13 for an illustration. Update T_H by removing every vertex that is not a part of G. Also, if the parent of s' in T_H is also in G, remove the edge between them in the updated tree. This results in a forest of at most $2w + 1$ subtrees of T_H that contains all vertices in G. This is because the roots of these subtrees can only be in the first and last bag of G, and at most $2w + 1$ vertices in G are located in these bags (note that s' is in both of the bags). Add edges from s' to the roots of all the subtrees. It is possible to add such edges because s' is connected to every vertex in G as it appears in all bags of G, and also, all vertices that share a bag are connected in H (because H is a w-path). The result would be a connected tree T_G, spanning G and rooted at s'.

We show broadcasting in T_G, starting from s', does not take more than $2w$ extra rounds than broadcasting in T_H, that is, $\boldsymbol{br}^*(G, s') \leq \boldsymbol{br}^*(H, s) + 2w$. This holds because s' can inform all roots of the subtrees first within $2w$ rounds. The remainder of broadcasting is conducted similarly in T_H and T_G. □

To establish a lower bound for broadcasting in any connected w-path graph G, we prove a slightly stronger result. Suppose we have k connected w-path graphs $H_1, H_2, \ldots, H_k$, each H_i having at least two vertices, a standard path decomposition of width w and a vertex $s_i \in H_i$ located at the left-most bag in its decomposition. We prove a lower bound for the total time for broadcasting in these graphs as follows. This is a generalization of our desired lower bound for the case of $k = 1$.

Lemma 5.4. *Consider a set of $k \geq 1$ vertex disjoint connected w-path graphs $\mathcal{H} = \{H_1, \ldots, H_k\}$. Suppose any H_i has a standard path decomposition D_i with ℓ_i bags, where a source vertex s_i is located in the leftmost bag of D_i. Let $L = \sum_{i=1}^{k} \ell_i$. Then we have $\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) \geq f(L, w)$, where $f(L, w) = 27^{-w}(w!)^{-2}L^{4^{-w}}$.*

Proof. Let n denote the total size (number of vertices) in all H_i's. We use an induction on n to prove the lemma. In the base of induction, we consider the cases where $n \leq 3w$, which implies $L \leq 3w$. Therefore, by the definition for $f(L, w)$, we have $f(L, w) \leq 1$. On the other hand, we have $\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) \geq 1$; this is because H_1 is formed by at least two vertices, and its broadcast time is at least 1. Thus, $f(L, w) \leq \boldsymbol{br}^*(H_1, s_1) \leq \sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i)$, and the base of induction holds.

In the induction step, we consider two cases, namely $k = 1$ and $k > 1$. First, suppose $k > 1$. Thus, each H_i has a size $|H_i| < n$. Therefore, we can apply the induction hypothesis on a set of graphs formed by only one H_i (for each $i \in [k]$). This gives $\boldsymbol{br}^*(H_i, s_i) \geq f(\ell_i, w)$. Summing over all i's and observing that the function $f(L, w)$ is sublinear in L, we can conclude

$$\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) \geq \sum_{i=1}^{k} f(\ell_i, w) \geq f(L, w).$$

Next, suppose $k = 1$. Consider the bags in the standard path decomposition D_1 containing any vertex v. By the definition of path decomposition, these bags form a consecutive set of bags in D_1; we refer to this set as the **span** of v and denote it by $\text{span}(v)$. Let v_m be the vertex with maximum span, i.e., $v_m = \arg\max_v |\text{span}(v)|$, and let $M = |\text{span}(v_m)| - 1$. For example, in Figure 13a, we have $span(a) = 2$, $v_m = d$, and $M = 3$. We analyze two cases based on the value of M.

Case 1: M is small. Suppose $M \leq \frac{L-1}{f(L,w)}$ and let u denote any vertex that only appears in the rightmost bag of D_1 (since D_1 is a standard path decomposition, at least one such vertex u exists). We claim the distance between the source s and u is large enough to establish the desired lower stated in the lemma. Let the shortest path between s and u be $\langle s(= u_0), u_1, \ldots, u_d(= u)\rangle$. For any $j \in [d]$ given that the span of u_j is at most M and u_j is connected to u_{j+1}, it must be that u_{j+1} appears for the first time in a bag within distance M (in D_1) of the bag where u_j was first introduced. It means the total number L of bags cannot be more than $dM + 1$ or $d \geq (L-1)/M$. Given that the distance between s and u in H_1 is an upper bound for the broadcast time, we can conclude

$$\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) = \boldsymbol{br}^*(H_1, s_1) \geq (L-1)/M \geq f(L, w).$$

The last inequality holds because of the assumption that $M \leq \frac{L-1}{f(L,w)}$.

Case 2: M is large. Suppose $M > \frac{L-1}{f(L,w)}$. Let B_s and B_e be the first and the last bag of D_1 in which v_m appears, respectively. Consider the subgraph G induced by all vertices that appear in any of the bags between B_s and B_e (inclusive of all vertices in B_s and B_e). Remove from G the vertex v_m; the result would be a graph $G' = \{G'_1, G'_2, \ldots, G'_r\}$, which is a set of $r(\geq 1)$ connected components which are possibly singletons (a graph with one vertex). Let c denote the number of such singletons ($c \geq 0$) and consider a graph resulting from removing these c singletons from G'. We denote the result as $\mathcal{H}' = \{H'_1, H'_2, \ldots, H'_{r-c}\}$, which is a set of vertex-disjoint connected w'-paths. Each H'_i has a standard path decomposition D'_i with ℓ'_i bags. Also, we consider an arbitrary s'_i for each H'_i that appears in the first bag of D'_i.

We have $w' \leq w - 1$ because v_m is removed from G to form G', and it has been present in all bags in the path decomposition of G'. Next, we consider two cases as follows based on the value of c.

- First, we establish the lemma when $c \geq f(L, w) + 2$. We note that each singleton component, except the two that may appear in B_s and B_e, finds v_m as their sole neighbor in H_1 (this holds because all vertices that share a bag in D_1 are connected). Therefore, the degree of v_m in any broadcast tree of H_1 is at least $c - 2 \geq f(L, w)$. We conclude that $\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) = \boldsymbol{br}^*(H_1, s_1) \geq f(L, w)$, and the lemma holds as desired.

- Next, we consider the case $c < f(L, w) + 2$. Since v_m appears in all bags from B_s to B_e in D_1, we can apply Lemma 5.3 to conclude that

$$\boldsymbol{br}^*(G, v_m) \leq \boldsymbol{br}^*(H_1, s_1) + 2w. \tag{13}$$

Therefore, if we can show $\boldsymbol{br}^*(G, v_m) \geq f(L, w) + 2w$, it follows that $\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) = \boldsymbol{br}^*(H_1, s_1) \geq \boldsymbol{br}^*(G, v_m) - 2w \geq f(L, w)$, which will establish the statement of the lemma as desired.

It remains to prove $\boldsymbol{br}^*(G, v_m) \geq f(L, w) + 2w$, for which we will use the induction hypothesis and Lemma 5.1 as follows. The number of the vertices in $\mathcal{H}'$ is at most $n - 1$; this is because $\mathcal{H}'$ misses v (along with possibly other vertices) compared to G, and the size of G is at most n. Let $L' = \sum_i \ell'_i$ be the total number of bags in $\mathcal{H}'$. We have $L' = M - c$ because every $H'_i \in \mathcal{H}'$ contains a consecutive set of bags of G, and c singleton components, each of which located in exactly one bag, are excluded. Therefore, we can use the induction hypothesis to write

$$\sum_i \boldsymbol{br}^*(H'_i, s'_i) \geq \sum_i f(\ell'_i, w - 1) \geq f(L', w - 1).$$

The last step is because function $f(\ell', w)$ is sublinear in ℓ' and $\sum_i \ell'_i = L'$. Moreover, if we apply Lemma 5.1 on G, when v_m is removed from G, we can conclude

$$\boldsymbol{br}^*(G, v_m)(2\boldsymbol{br}^*(G, v_m) + 1) \geq \sum_{j \in [p+c]} \boldsymbol{br}^*(G'_j, s'_j) = \sum_{i \in [r-c]} \boldsymbol{br}^*(H'_i, s'_i). \tag{14}$$

The last inequality holds because broadcasting in each of the singleton graphs that are removed from G' to form $\mathcal{H}'$ takes 0 round. Therefore, we can write

$$\begin{aligned}
3\boldsymbol{br}^*(G, v_m)^2 \geq& \boldsymbol{br}^*(G, v_m)(2\boldsymbol{br}^*(G, v_m) + 1) && \boldsymbol{br}^*(G, v_m) \geq 1 \\
=& \sum_{i \in [r-c]} \boldsymbol{br}^*(H'_i, s'_i) && \text{from (14)} \\
\implies \boldsymbol{br}^*(G, v_m) \geq& \Big(\sum_{i \in [r-c]} \boldsymbol{br}^*(H'_i, s'_i)/3 \Big)^{1/2} && \\
\geq& \big(f(L', w - 1)/3 \big)^{1/2}. && \text{by the induction hypothesis}
\end{aligned}$$

We claim that $L' \geq \sqrt{L}$. Assuming this claim holds, continuing from the above inequality, we can write

$$\begin{aligned}
\boldsymbol{br}^*(G, v_m) \geq& \big(f(L', w - 1)/3 \big)^{1/2} \geq \big(f(\sqrt{L}, w - 1)/3 \big)^{1/2} && \\
\geq& \big(27^{-w+1} L^{4^{-w+1}/2} ((w - 1)!)^{-2}/3 \big)^{1/2} && \text{by definition of } f(.) \\
\geq& 3(27)^{-w} L^{4^{-w}} (w!)^{-2} w && \\
\geq& 3w f(L, w) && \text{by definition of } f(.) \\
\geq& f(L, w) + 2w. && \text{because } f(L, w) \geq 1
\end{aligned}$$

Recall that $\boldsymbol{br}^*(H_1, s_1) \geq \boldsymbol{br}^*(G, v_m) - 2w$ by Inequality (13) and combining this with the above inequality, we can conclude $\sum_{i=1}^{k} \boldsymbol{br}^*(H_i, s_i) = \boldsymbol{br}^*(H_1, s_1) \geq f(\ell, w)$. Now, it remains to prove that $L' \geq \sqrt{L}$.

$$\begin{aligned}
L' &= M - c \\
&> M - f(L,w) - 2 && c \text{ is assumed less than } f(L,w)+2 \\
&> \frac{L-1}{f(L,w)} - f(L,w) - 2. && M \text{ is assumed larger than } \frac{L-1}{f(L,w)}
\end{aligned}$$

Therefore, to prove $L' \geq \sqrt{L}$ it suffices to prove $\frac{L-1}{f(L,w)} - f(L,w) - 2 \geq \sqrt{L}$, which we establish as follows.

$$\begin{aligned}
&\frac{L-1}{f(L,w)} - f(L,w) - 2 \geq \sqrt{L} \\
\Longleftrightarrow\ & L - 1 \geq \sqrt{L} f(L,w) + f(L,w)^2 + 2f(L,w) \\
\Longleftarrow\ & L - 1 \geq 2L^{3/4}/27 + L^{1/2}/27^2 + 2L^{1/4}/27 && \text{by definition of } f(L,w) \\
\Longleftarrow\ & L - 1 \geq L^{3/4}/3.
\end{aligned}$$

The last inequality is true for any $L \geq 2$.

□

The following lemma generalizes Lemma 5.4 (applied with $k = 1$) to the case where the source is not necessarily located in the first bag of the path decomposition.

Lemma 5.5. *Let G be any connected graph and let s be any vertex of G. Suppose G has a standard path decomposition of width w formed by L bags. Then, we have $\boldsymbol{br}^*(G,s) \geq f(L-1, w+1)$, where $f(L,w) = 27^{-w}(w!)^{-2}L^{4^{-w}}$.*

Proof. Let $D = \langle B_1, \ldots, B_L \rangle$ be the standard path decomposition of G of width w, and B_k be the leftmost bag that contains s in D. Add s to any bag B_x of D for $x \in [k-1]$. The result would be a valid path decomposition $D' = \langle B'_1, \ldots, B'_{L'} \rangle$ of width w', which is not larger than $w+1$. To keep the path decomposition in the standard form, we may need to merge B_k with B_{k-1}. Regardless, L' is at least $L-1$.

We then add edges between s and all vertices of G that appear in a bag B_i where $i \leq k$; the result would be a new graph G^+ that is a $(w+1)$-path (recall that adding edges will not increase the broadcast time). Now, we have a Telephone Broadcasting instance (G^+, s), where G^+ is a $(w+1)$-path and the source is located on B_1. Therefore, we can apply Lemma 5.4 on a single component $H_1 = G^+$ ($k = 1$), to conclude we have $\boldsymbol{br}^*(G,s) \geq \boldsymbol{br}^*(G^+, s) \geq f(L', w') \geq f(L-1, w+1)$. □

We are now ready to prove the main result of this section.

Proof of Theorem 5.2: Let (G,s) be an instance of Telephone Broadcasting, where G is a graph of pathwidth w. Consider a standard path decomposition D of G, where each bag contains at most $w+1$ vertices. Let L be the number of bags in D and note that $L \geq \frac{n}{w+1}$. This holds because each vertex has to be present in at least one bag. Combining this inequality with Lemma 5.5, we will obtain

$$\boldsymbol{br}^*(G,s) \geq f(L-1, w+1) = 27^{-(w+1)}((w+1)!)^{-2}\Big(\frac{n}{2(w+1)} - 1\Big)^{4^{-(w+1)}}. \tag{15}$$

Therefore, assuming $w = \mathcal{O}(1)$, we can write $\boldsymbol{br}^*(G, s) = \Omega(n^{4^{-(w+1)}})$. The approximation ratio of the algorithm introduced by Elkin and Kortsarz [7] is given by $\mathcal{O}\left(\log n / \log \boldsymbol{br}^*(G, s)\right)$. Using Inequality (15), the approximation factor would be

$$\mathcal{O}\Big(\frac{\log n}{\log n^{4^{-(w+1)}}}\Big) = \mathcal{O}(4^w),$$

which completes the proof. □

6 Concluding Remarks

In this paper, we resolved an open problem by proving the NP-completeness of TELEPHONE BROADCASTING for cactus graphs as well as graphs of pathwidth 2. We also established a 2-approximation algorithm for cactus graphs and a constant-factor approximation algorithm for graphs of bounded pathwidth. A possible direction for future work is to improve the approximation factor for graphs of bounded pathwidth or cactus graphs. In particular, it remains an open question whether Polynomial Time Approximation Schemes (PTASs) exist for these graph classes. A major open problem in this domain is determining whether a constant-factor approximation exists for general graphs. While progress on this question has been slow, the algorithmic ideas developed in this work may be applicable to broadcasting in other families of sparse graphs.

Acknowledgement

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC)[funding reference number DGECR-2018-00059].

References

[1] Boaz Ben-Moshe, Amit Dvir, Michael Segal, and Arie Tamir. Centdian computation in cactus graphs. *Journal of Graph Algorithms and Applications*, 16(2):199–224, 2012.

[2] Puspal Bhabak and Hovhannes A. Harutyunyan. Constant approximation for broadcasting in k-cycle graph. In *Proceedings of first Conference of Algorithms and Discrete Applied Mathematics*, volume 8959, pages 21–32, 2015.

[3] Maja Čevnik and Janez Žerovnik. Broadcasting on cactus graphs. *Journal of Combinatorial optimization*, 33:292–316, 2017.

[4] Peter Damaschke. A linear-time optimal broadcasting algorithm in stars of cliques. *Journal of Graph Algorithms and Applications*, 28(1):385–388, 2024.

[5] Anne-Laure Ehresmann. Approximation algorithms for broadcasting in flower graphs. Master's thesis, Concordia University, 2021.

[6] Michael Elkin and Guy Kortsarz. Combinatorial logarithmic approximation algorithm for directed telephone broadcast problem. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 438–447, 2002.

[7] Michael Elkin and Guy Kortsarz. Sublogarithmic approximation for telephone multicast. *Journal of Computer and System Sciences*, 72(4):648–659, 2006.

[8] Fraigniaud and Mitjana. Polynomial-time algorithms for minimum-time broadcast in trees. *Theory of Computing Systems*, 35(6):641–665, 2002.

[9] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. 1979.

[10] Saber Gholami, Hovhannes A Harutyunyan, and Edward Maraachlian. Optimal broadcasting in fully connected trees. *Journal of Interconnection Networks*, 23(01):2150037, 2023.

[11] Michelangelo Grigni and David Peleg. Tight bounds on minimum broadcast networks. *SIAM Journal on Discrete Mathematics*, 4(2):207–222, 1991.

[12] Hovhannes Harutyunyan and Edward Maraachlian. Linear algorithm for broadcasting in unicyclic graphs. In *Proceedings of the 13th Annual International Conference on Computing and Combinatorics*, pages 372–382, 2007.

[13] Hovhannes Harutyunyan, George Laza, and Edward Maraachlian. Broadcasting in necklace graphs. In *Proceedings of the 2nd Canadian Conference on Computer Science and Software Engineering*, pages 253–256, 2009.

[14] Hovhannes A Harutyunyan and Narek Hovhannisyan. Improved approximation for broadcasting in k-path graphs. In *International Conference on Combinatorial Optimization and Applications*, pages 111–122, 2023.

[15] Hovhannes A Harutyunyan and Arthur L Liestman. Improved upper and lower bounds for k-broadcasting. *Networks: An International Journal*, 37(2):94–101, 2001.

[16] Hovhannes A Harutyunyan and Edward Maraachlian. On broadcasting in unicyclic graphs. *Journal of combinatorial optimization*, 16:307–322, 2008.

[17] Hovhannes A Harutyunyan, Narek Hovhannisyan, and Edward Maraachlian. Broadcasting in chains of rings. In *Proceedings of the Fourteenth International Conference on Ubiquitous and Future Networks*, pages 506–511, 2023.

[18] Sandra M Hedetniemi, Stephen T Hedetniemi, and Arthur L Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18(4):319–349, 1988.

[19] Guy Kortsarz and David Peleg. Approximation algorithms for minimum-time broadcast. *SIAM Journal on Discrete Mathematics*, 8(3):401–427, 1995.

[20] Arthur L. Liestman and Joseph G. Peters. Broadcast networks of bounded degree. *SIAM Journal on Discrete Mathematics*, 1(4):531–540, 1988.

[21] Andrzej Proskurowski and Jan Arne Telle. Classes of graphs with restricted interval models. *Discret. Math. Theor. Comput. Sci.*, 3(4):167–176, 1999.

[22] Neil Robertson and Paul D Seymour. Graph minors. i. excluding a forest. *Journal of Combinatorial Theory, Series B*, 35(1):39–61, 1983.

[23] Weiping Shang, Pengjun Wan, and Xiaodong Hu. Approximation algorithms for minimum broadcast schedule problem in wireless sensor networks. *Frontiers of Mathematics in China*, 5:75–87, 2010.

[24] Peter J. Slater, Ernest J. Cockayne, and Stephen T. Hedetniemi. Information dissemination in trees. *SIAM Journal on Computing*, 10(4):692–701, 1981.

[25] Elena Stöhr. Broadcasting in the butterfly network. *Information Processing Letters*, 39(1): 41–43, 1991.

[26] Prafullkumar Tale. Double exponential lower bound for telephone broadcast. *arXiv preprint arXiv:2403.03501*, 2024.

[27] Craig A Tovey. A simplified np-complete satisfiability problem. *Discrete applied mathematics*, 8(1):85–89, 1984.