

# Analyzing the Optimal Neighborhood: Algorithms for Budgeted and Partial Connected Dominating Set Problems\*

Samir Khuller<sup>†</sup>      Manish Purohit<sup>‡</sup>      Kanthi K. Sarpatwar<sup>§</sup>

## Abstract

We study *partial* and *budgeted* versions of the well studied connected dominating set problem. In the partial connected dominating set problem (PCDS), we are given an undirected graph  $G = (V, E)$  and an integer  $n'$ , and the goal is to find a minimum subset of vertices that induces a connected subgraph of  $G$  and dominates at least  $n'$  vertices. We obtain the first polynomial time algorithm with an  $O(\ln \Delta)$  approximation factor for this problem, thereby significantly extending the results of Guha and Khuller (Algorithmica, Vol. 20(4), Pages 374-387, 1998) for the connected dominating set problem. We note that none of the methods developed earlier can be applied directly to solve this problem. In the budgeted connected dominating set problem (BCDS), there is a budget on the number of vertices we can select, and the goal is to dominate as many vertices as possible. We obtain a  $\frac{1}{13}(1 - \frac{1}{e})$  approximation algorithm for this problem. Finally, we show that our techniques extend to a more general setting where the profit function associated with a subset of vertices is a “special” submodular function. This generalization captures the connected dominating set problem with capacities and/or weighted profits as special cases. This implies a  $O(\ln q)$  approximation (where  $q$  denotes the quota) and an  $O(1)$  approximation algorithms for the partial and budgeted versions of these problems. While the algorithms are simple, the results make a surprising use of the greedy set cover framework in defining a useful profit function.

## 1 Introduction

A *connected dominating set* (CDS) in a graph is a dominating set that induces a connected subgraph. The CDS problem, which seeks to find the minimum such set, has been widely studied [49, 23, 32, 51, 20,

27, 22, 42, 16, 17] starting from the work of Guha and Khuller [32]. The CDS problem is NP-hard and thus the literature has focused on the development of fast polynomial time approximation algorithms. For general graphs, Guha and Khuller [32] propose an algorithm with a  $\ln \Delta + 3$  approximation factor, where  $\Delta$  is the maximum degree of any vertex. Better approximation algorithms are known in special classes of graphs. For the case of planar [21] or geometric unit disk graphs [18] polynomial time approximation schemes (PTAS) are known. This problem has also been extensively studied in the distributed setting [49, 23]. Not surprisingly, CDS problem in general graphs is at least as hard to approximate as the set cover problem for which a hardness result of  $(1 - \epsilon) \log n$  (unless  $NP \subseteq DTIME(n^{O(\log \log n)})$ ) follows by the work of Feige [25].

CDS has become an extremely popular topic, for example, the recent book by Du and Wan [22] focuses on the study of ad hoc wireless networks as CDSS provide a platform for routing on such networks. In these ad hoc wireless networks, a CDS can act as a virtual backbone so that only nodes belonging to the CDS are responsible for packet forwarding and routing. Minimizing the number of nodes in the virtual backbone leads to increased network lifetime, and lesser bandwidth usage due to control packets, and hence the CDS problem has been extensively studied and applied to create such virtual backbones.

One shortcoming of using a CDS as a virtual backbone is that a few distant clients (outliers) can have the undesirable effect of increasing the size of the CDS without improving the quality of service to a majority of the clients. In such scenarios, it is often desirable to obtain a much smaller backbone that provides services to, say, (at least) 90% of the clients. Liu and Liang [42] study this problem of finding a minimum *partial connected dominating set* in wireless sensor networks (geometric disk graphs) and provide heuristics (without guarantees) for the same.

A complementary problem is the budgeted CDS problem where we have a budget of  $k$  nodes, and we wish

\*The work is supported by NSF grants CCF 1217890 and CCF 0937865.

<sup>†</sup>Computer Science Department, University of Maryland, College Park. E-mail: samir@cs.umd.edu

<sup>‡</sup>Computer Science Department, University of Maryland, College Park. E-mail: manishp@cs.umd.edu

<sup>§</sup>Computer Science Department, University of Maryland, College Park. E-mail: kasarpa@cs.umd.edu

to find a connected subset of  $k$  nodes which dominate as many vertices as possible. Budgeted domination has been studied in sensor networks where bandwidth constraints limit the number of sensors we can choose and the objective is to maximize the number of targets covered [16, 17].

Another application arises in the context of social networks. Consider a social network where vertices of the network correspond to people and an edge joins two vertices if the corresponding people influence each other. Avrachenkov et al. [2] consider the problem of choosing  $k$  connected vertices having maximum total influence in a social network using local information only (i.e., the neighborhood of a vertex is revealed only after the vertex is bought) and provide heuristics (without guarantees) for the same. Borgs et al. [9] show that no local algorithm for the partial dominating set problem can provide an approximation guarantee better than  $O(\sqrt{n})$ . As the influence of a set of vertices is simply the number of dominated vertices, these problems are exactly the budgeted and partial connected dominating set problems with the additional restriction of local only information.

Budgeted versions of set cover (known as max-coverage)<sup>1</sup> are well understood and the standard greedy algorithm is known to give the optimal  $1 - \frac{1}{e}$  approximation [45]. Khuller et al. [40] give a  $(1 - \frac{1}{e})$  approximation algorithm for a generalized version with costs on sets. In addition, we may consider a partial version of the set cover problem, also known as partial covering, in which we wish to pick the minimum number of sets to cover a pre-specified number of elements. Kearns [39] first showed that greedy gives a  $2H_n + 3$  approximation guarantee (where  $n$  is the ground set cardinality and  $H_n$  is the  $n^{\text{th}}$  harmonic number), which was improved by Slavík [47] to obtain a guarantee of  $\min(H_{n'}, H_\Delta)$  (where  $n'$  is the minimum coverage required and  $\Delta$  is the maximum size of any set). Wolsey [50] considered the more general *submodular cover* problem and showed that the simple greedy delivers a best possible  $\ln n$  approximation. For the case where each element belongs to at most  $f$  (called the *frequency*) different sets, Gandhi et al. [26], using a primal-dual algorithm, and Bar-Yehuda [4], using the local-ratio technique, achieve an  $f$ -approximation guarantee.

Unfortunately for both the budgeted and partial versions of the CDS problem, greedy approaches based on prior methods fail. The fundamental reason is that while the greedy algorithm works well as a method for

rapidly “covering” nodes, the cost to connect different chosen nodes can be extremely high if the chosen nodes are far apart. On the other hand if we try to maintain a connected subset, then we cannot necessarily select nodes from dense regions of the graph. In fact, none of the approaches in the work by Guha and Khuller [32] appear to extend to these versions.

Partial and budgeted optimization problems have been extensively studied in the literature. Most of these problems, with the exception of partial and budgeted set cover, required significantly different techniques and ideas from the corresponding “complete” versions. We will now cite several such problems.

The best example is the minimum spanning tree problem, which is well known to be polynomial time solvable. However the partial version of this problem where we look for a minimum cost tree which spans at least  $k$  vertices is NP-hard [46]. A series of approximation techniques [1, 3, 8, 28] finally resulted in a 2-approximation [29] for the problem.

Partial versions of several classic location problems like  $k$ -center and  $k$ -median have required new techniques as well. The partial  $k$ -center problem, which is also called the outlier  $k$ -center problem or the robust  $k$ -center problem, requires us to minimize the maximum distance to the “best”  $n'$  nodes (while the complete version requires us to consider all the nodes) to the centers. Charikar et al. [13] gave a 3-approximation algorithm whose analysis was significantly different from the classic  $k$ -center 2-approximation algorithm [31, 36]. Chen [15] gives a constant approximation for outlier  $k$ -median problem, while Charikar et al. [13] gave a 4-approximation for the outlier uncapacitated facility location problem.

Several other optimization problems need special techniques to tackle the corresponding partial versions. Notable examples of such problems, include *partial vertex cover* [48, 5, 43, 26, 10, 35], quota Steiner tree problems [37], budgeted and partial node weighted Steiner tree problems [44, 7], and scheduling with outliers [12, 33]. We end this subsection by noting that partial versions of some optimization problems are completely inapproximable even though, the corresponding complete version has a small constant approximation algorithm. The best example of this is the *robust subset resource replication* problem studied by Khuller et al. [41].

**1.1 Other Related Work** In the *group Steiner tree problem*, we are given a graph  $G = (V, E)$ , an associated cost function  $c : E \rightarrow \mathbb{R}^+ \cup \{0\}$ , and a collection of groups of vertices  $g_1, g_2, \dots, g_k$ . The goal is to find a minimum cost tree that contains at least one vertex from

<sup>1</sup>Here instead of finding the smallest sub-collection of sets to cover a given set of elements, we fix a budget on number of sets we wish to pick with the objective of maximizing the number of covered elements.

each group. It can be observed that the connected dominating set problem reduces to the group Steiner tree problem by creating a group for every vertex containing the neighborhood of that vertex. Garg et al. [30] obtain a  $O(\log(\max_{i \in [k]} |g_i|) \log k)$  approximation for this problem, in the special case when the graph is a tree. Using a decomposition result due to Bartal [6], Garg et al. [30] extend the tree result to obtain a  $O(\log^3 n \log k)$  approximation algorithm for arbitrary graphs. Fakcharoenphol et al. [24] improve Bartal's decomposition result, consequently obtaining a  $O(\log^2 n \log k)$  approximation for the group Steiner tree problem in arbitrary graphs. Halperin et al. [34] note that Garg et al. [30]'s algorithm also gives a  $O(\log(\max_{i \in [k]} |g_i|))$  approximation for the budgeted group Steiner tree problem on trees. They also show a  $\log^{2-\epsilon} k$  hardness of approximation for the (partial) group Steiner problem and a  $\log^{1-\epsilon} k$  hardness of approximation for the budgeted version. Chekuri et al. [14] gave a combinatorial algorithm for the group Steiner tree problem on trees, with an approximation guarantee of  $O((\log \sum_i |g_i|)^{1+\epsilon} \log k)$ . Calinescu and Zelikovsky [11] extended Chekuri et al. [14]'s result to the more general problem of *polymatroid Steiner tree*.

**1.2 Our Contributions** Our results can be summarized as follows

- In Section 4, we obtain the first  $O(\ln \Delta)$  approximation algorithm for the PCDS problem. To be precise, our approximation guarantee is  $4 \ln \Delta + 2 + o(1)$ , where  $\Delta$  is the maximum degree.
- In Section 5, we obtain a  $\frac{1}{13}(1 - \frac{1}{e})$ -approximation algorithm for the BCDS problem. This is the first constant approximation known for BCDS.
- In Section 6, we generalize the above problems to a special kind of submodular optimization problem (to be defined later), which has *capacitated connected dominating set* problem and *weighted profit connected dominating set* problem as special cases. Again, we obtain  $O(\ln q)$  and  $\frac{1}{13}(1 - \frac{1}{e})$  approximation algorithms for the partial and budgeted version of this problem respectively where  $q$  denotes the quota for the partial version.

## 2 Preliminaries

We now proceed to formally define the problems we consider in this paper.

**PARTIAL CONNECTED DOMINATING SET PROBLEM (PCDS).** Given an undirected graph  $G = (V, E)$ , and an integer (quota)  $n'$ , find a minimum size subset  $S \subseteq V$  of vertices such that the graph induced by  $S$  is connected, and  $S$  dominates at least  $n'$  vertices.

**BUDGETED CONNECTED DOMINATING SET PROBLEM (BCDS).** Given an undirected graph  $G = (V, E)$ , and an integer (budget)  $k$ , find a subset  $S \subseteq V$  of at most  $k$  vertices such that the graph induced by  $S$  is connected, and the number of vertices dominated by  $S$  is maximized.

Before defining the remaining problems, we introduce the notion of *special submodularity*.

**SPECIAL SUBMODULAR FUNCTION.** Let  $G = (V, E)$  be an arbitrary graph. A function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup \{0\}$ , is said to have the *special submodular* property if it satisfies the following-

- $f$  is submodular. That is  $f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad \forall A, B, v$  such that  $A \subseteq B \subseteq V$ .
- $f_A(X) = f_{A \cup B}(X)$ , if  $N(X) \cap N(B) = \emptyset \quad \forall X, A, B \subseteq V$ .

where  $f_A(X) = f(A \cup X) - f(A)$  is the marginal profit of  $X$  given  $A$  and  $N(X)$  denotes the neighborhood of  $X$ , including  $X$  itself.

We now define the generalized versions of PCDS and BCDS.

**PARTIAL GENERALIZED CONNECTED DOMINATING SET PROBLEM (PGCDS).** Given a graph  $G = (V, E)$ , an integer (quota)  $q$ , and a monotone *special submodular* profit function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup \{0\}$ , find a subset  $S \subseteq V$  of minimum size, such that  $f(S) \geq q$  and  $S$  induces a connected subgraph in  $G$ .

**BUDGETED GENERALIZED CONNECTED DOMINATING SET PROBLEM (BGCDS).** Given a graph  $G = (V, E)$ , a budget  $k$ , and a monotone *special submodular* profit function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup \{0\}$ , find a subset  $S \subseteq V$  which maximizes  $f(S)$  such that  $|S| \leq k$  and  $S$  induces a connected subgraph of  $G$ .

These problems capture the following variants of partial and budgeted connected dominating set problems.

1. **WEIGHTED PROFIT CONNECTED DOMINATING SET.** In this variant, each vertex has an arbitrary profit which is obtained if it is dominated by some chosen vertex.
2. **CAPACITATED CONNECTED DOMINATING SET.** In this variant, each vertex has a capacity which is the number of vertices it can dominate.

For all of our algorithms we will be using an algorithm for the Quota Steiner Tree problem (QST) as a subroutine. We now define the QST problem and mention relevant results.

**QUOTA STEINER TREE PROBLEM (QST).** Given an undirected graph  $G = (V, E)$ , a profit function  $p : V \rightarrow \mathbb{Z}^+ \cup \{0\}$  on the vertices, a cost function  $c : E \rightarrow$

$Z^+ \cup \{0\}$  on the edges, and an integer (quota)  $q$ , find a subtree  $T$  that minimizes  $\sum_{e \in E(T)} c(e)$ , subject to  $\sum_{v \in V(T)} p(v) \geq q$ .

Johnson et al. [37] studied the QST problem and showed that an  $\alpha$ -approximation algorithm for the  $k$ -MST problem can be adapted to obtain an  $\alpha$ -approximation algorithm for the Quota Steiner Tree problem. Using this result along with the 2-approximation for  $k$ -MST by Garg [29], gives us the following theorem.

**THEOREM 2.1.** ([37, 29]) *There is a 2-approximation algorithm for the Quota Steiner Tree Problem.*

### 3 Shortcomings of Prior Approaches.

We now describe the three approaches taken by Guha and Khuller [32] to solve the CDS problem and show why none of these approaches extend directly for the budgeted and partial coverage variants.

*Algorithm 1.* The first algorithm is a “one step look-ahead” greedy algorithm where they iteratively grow a tree by selecting a *pair* of vertices that together cover the most number of previously uncovered vertices. Figure 1 shows a bad instance on which a  $c$ -step look-ahead greedy algorithm fails for the BCDS and PCDS. The instance contains  $k$  “spiders” whose heads (vertices with degree  $> 2$ ) are connected by paths of length  $c+1$ . The spider heads are the only vertices that offer profit greater than 3. We show that on this graph, there are BCDS and PCDS instances that can perform very poorly. Consider a BCDS instance on the graph, with a budget  $k + (c+1)(k-1)$ . Clearly the optimal solution picks the path connecting all the spider heads, so that the total coverage is  $(M+1)k + (c+1)(k-1)$ . On the other hand, the  $c$ -step look-ahead greedy algorithm, might get stuck inside one of the spiders and may end up selecting as many as  $M+1$  vertices from it. This is because, despite the look-ahead capability of the algorithm, the spider legs will become indistinguishable from the optimal path. For a sufficiently large value of  $M$ , the  $c$ -step look-ahead algorithm might use up all its budget on a single spider, there by obtaining a coverage of  $O(M+k)$ . Thus in the worst case the look-ahead greedy algorithm could have a  $\Omega(k)$  approximation guarantee. Using a similar argument, we can show that, for the PCDS instance on the graph with quota  $Mk$ , the approximation guarantee could be  $\Omega(M)$ .

*Algorithm 2.* The second algorithm is to find a dominating set  $D$  and run a Steiner tree algorithm with the vertices in  $D$  as terminals. Since the optimal connected dominating set, by definition, is a tree that dominates  $D$ , we can show that there exists a Steiner tree of low cost with the set  $D$  as terminals. Using a

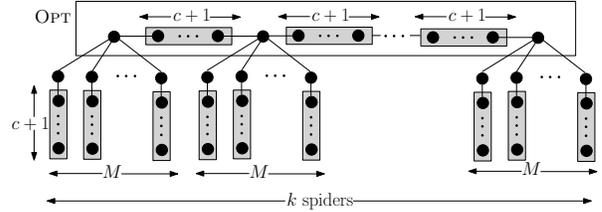


Figure 1: A bad example for the  $c$ -step look-ahead greedy algorithm

constant factor approximation algorithm for the Steiner tree problem, we obtain a  $O(\ln n)$  approximation for the connected dominating set. However, for the partial and budgeted versions, the optimal solution does not dominate all vertices and hence it’s not possible to bound the cost of the Steiner tree in terms of the optimal solution.

*Algorithm 3.* The final algorithm builds unconnected components greedily and owing to the fact that every vertex has to be dominated, makes sure that the constructed components be connected cheaply. Again this approach fails in the partial and budgeted case because the components created when we have dominated a specified number of vertices could be far apart.

### 4 Partial Connected Dominating Set

In this section, we consider the partial connected dominating set (PCDS) problem and give a  $4 \ln \Delta + 2 + o(1)$ -approximation algorithm for the same.

**4.1 Algorithm** We now give a high level overview of the algorithm. The algorithm itself is very simple but to show that it is indeed a  $O(\log \Delta)$  approximation requires non-trivial analysis. The algorithm proceeds in the following manner. We first run a simple greedy algorithm to find a (not necessarily connected) dominating set  $D$ . In each iteration, the greedy algorithm chooses a vertex that dominates the maximum number of previously undominated vertices. We call this number the “profit” associated with the chosen vertex. Given this profit function on the nodes, we now apply a 2-approximation algorithm for the Quota Steiner Tree (QST) problem, with quota of  $n'$  to obtain a connected solution<sup>2</sup>.

This is a little surprising, since the profit function depends on the choices made by the greedy algorithm

<sup>2</sup>Note that we could have simply defined each node’s profit as the number of vertices it can dominate and then try to connect nodes using the algorithm for the QST problem, however in this setting there could be a set of high profit nodes that get chosen, but since they all dominate the *same* subset of nodes, we do not actually gain a profit of  $n'$ .

in the first phase. However, we can show that there is a subset of vertices  $D' \subseteq D$ , of cardinality at most  $|\text{OPT}| \ln \Delta + 1$  whose profits sum up to at least  $n'$  where  $|\text{OPT}|$  is the size of the optimum solution of the PCDS instance. Furthermore the vertices in  $D'$  can be connected with additional  $(\ln \Delta + 1)|\text{OPT}| + 1$  vertices. Thus, if we could find the smallest tree with total profit at least  $n'$ , such a tree would cost (number of edges in the tree) no more than  $(2 \ln \Delta + 1)|\text{OPT}| + 1$ . This is a special case of the QST problem (with unit edge costs) and hence we can apply Theorem 2.1 to obtain a tree of size (cost) no more than  $2((2 \ln \Delta + 1)|\text{OPT}| + 1) = (4 \ln \Delta + 2)|\text{OPT}| + 2$ . Thus, we obtain a  $(4 \ln \Delta + 2 + o(1))$ -approximate solution for the PCDS problem.

---

**ALGORITHM 4.1. GREEDY PROFIT LABELING ALGORITHM FOR PCDS.**

**Input:** Graph  $G = (V, E)$  and  $n' \in \mathbb{Z}^+ \cup \{0\}$ .  
**Output:** Tree  $T$  with at least  $n'$  Coverage.

- 1: Compute the greedy dominating set  $D$  and the corresponding profit function  $p : V \rightarrow \mathbb{N}$  using the Algorithm 4.2.
  - 2: Use the 2-approximation for QST problem [37] on the instance  $(G, p)$  to obtain a tree  $T$  with profit at least  $n'$ .
- 

**ALGORITHM 4.2. GREEDY DOMINATING SET.**

**Input:** Graph  $G = (V, E)$ .  
**Output:** Dominating Set  $D$  and profit function  $p : V \rightarrow \mathbb{Z}^+ \cup \{0\}$ .

- 1:  $D \leftarrow \phi$
  - 2:  $U \leftarrow V$
  - 3: **for all**  $v \in V$  **do**
  - 4:      $p(v) \leftarrow 0$ ;
  - 5: **end for**
  - 6: **while**  $U \neq \phi$  **do**
  - 7:      $v \leftarrow \arg \max_{v \in V \setminus D} |N_U(v)| \quad \triangleright N_U(v)$  is the set of neighbors of  $v$ , including itself, in the set  $U$
  - 8:      $C_v \leftarrow N_U(v)$
  - 9:      $p(v) \leftarrow |C_v|$
  - 10:      $U \leftarrow U \setminus N_U(v)$
  - 11:      $D \leftarrow D \cup \{v\}$
  - 12: **end while**
- 

**4.2 Analysis** We first introduce some required notation.

**Notation:** For every vertex  $v \in D$  that is chosen by the greedy algorithm, let  $C_v$  denote the set of *new*

vertices that  $v$  dominates i.e., we have  $p(v) = |C_v|$ . We say that  $v$  “covers” a vertex  $w$  if and only if  $w \in C_v$ . For the sake of analysis, we partition the vertices of the graph  $G$  into layers. Let  $L_1 = \text{OPT}$  be the vertices in an optimal solution for the PCDS instance,  $L_2$  be the set of vertices that are not in  $L_1$  and have at least one neighbor in  $L_1$ , and  $R = V \setminus \{L_1 \cup L_2\}$  be the remaining vertices. Let  $L_3$  be the subset of vertices of  $R$  that have a neighbor in  $L_2$ . Furthermore let  $L'_i = D \cap L_i, 1 \leq i \leq 3$  where  $D$  is the dominating set chosen by the greedy algorithm. Figure 2 clarifies this notation regarding the layers  $L_i$ .

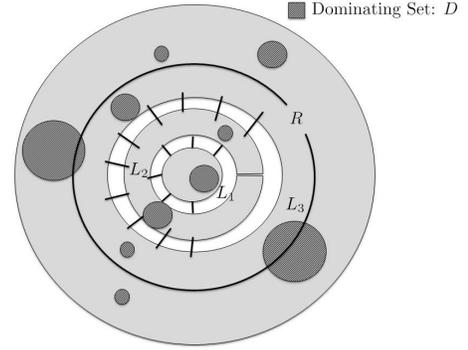


Figure 2: Pictorial Representation of Different Layers. (a)  $L_1$  is an optimal solution (b)  $L_2$  is set of the vertices adjacent to  $L_1$  (c)  $L_3$  is the subsequent layer (d)  $R$  is the set of all vertices other than  $L_1 \cup L_2$  and (e)  $L'_i = L_i \cap D$ .

We first show the following.

**LEMMA 4.1.** *There is a subset  $D' \subseteq L'_1 \cup L'_2 \cup L'_3$  such that  $|D'| \leq |\text{OPT}| \ln \Delta + 1$  and the total profit of vertices in  $D'$  is at least  $n'$ , i.e.  $\sum_{v \in D'} p(v) \geq n'$ .*

*Proof.* Let  $L'_1 \cup L'_2 \cup L'_3 = \{v_1, v_2, \dots, v_l\}$  where the vertices are arranged according to the order in which they were selected by the greedy algorithm. Since all vertices in  $L_1 \cup L_2$  are dominated by  $L'_1 \cup L'_2 \cup L'_3$ , we have  $\sum_{i=1}^l p(v_i) \geq |L_1 \cup L_2| \geq n'$  where the second inequality follows from the fact that  $L_1$  is a feasible solution (in fact optimal feasible solution). Choose  $t$  such that  $\sum_{i=1}^t p(v_i) < n'$  and  $\sum_{i=1}^{t+1} p(v_i) \geq n'$ . Let  $\mathcal{S} = \{v_1, v_2, \dots, v_t\}$  denote the set of the first  $t$  vertices chosen from the set  $L'_1 \cup L'_2 \cup L'_3$ . We now show that  $|\mathcal{S}| = t \leq |\text{OPT}| \ln \Delta$  and hence  $D' = \mathcal{S} \cup \{v_{t+1}\}$  satisfies the requirements of the claim.

Let  $C_{12}$  be the set of vertices in  $L_1 \cup L_2$  that are covered by  $\mathcal{S}$  in the original greedy step i.e.,  $C_{12} = \cup_{v \in \mathcal{S}} \{C_v \cap (L_1 \cup L_2)\}$ . Let  $UC_{12} = (L_1 \cup L_2) \setminus C_{12}$  be the vertices in  $L_1 \cup L_2$  that are not covered by  $\mathcal{S}$ . Similarly define  $C_R = \cup_{v \in \mathcal{S}} \{C_v \cap R\}$  as the set of vertices in  $R$

covered by  $\mathcal{S}$  (as per the greedy step). Then, we have that  $|C_R| + |C_{12}| < n' \leq |L_1 \cup L_2| = |C_{12}| + |UC_{12}|$ , where the first inequality follows from the definition of  $\mathcal{S}$ . Therefore we have  $|C_R| < |UC_{12}|$ .

We can thus assign every vertex in  $C_R$  to a unique vertex in  $UC_{12}$ , i.e. let  $I : C_R \rightarrow UC_{12}$  denote a one to one function from  $C_R$  to  $UC_{12}$ . In the subsequent charging argument, any cost that we charge to a vertex  $x \in C_R$  is transferred to its assigned vertex  $I(x) \in UC_{12}$ . Hence, after this charge transfer, only vertices in  $L_1 \cup L_2$  will be charged. We will now use a charging argument to show that  $|\mathcal{S}| \leq |\text{OPT}| \ln \Delta$ .

Consider a vertex  $u \in \mathcal{S}$ . We recall that  $C_u$  is the set of vertices covered for the first time by  $u$  in the greedy step. We assign every  $w \in C_u$  a charge  $\rho(w) = \frac{1}{|C_u|}$ . It is clear that the total charge on all vertices is equal to the size of  $\mathcal{S}$ . As described above, the charge of a vertex in  $w \in R$  is transferred to its mapped vertex in  $I(w) \in UC_{12}$ . Let  $v$  be a vertex in the optimal solution set  $L_1$ . We denote the set of neighbors of  $v$ , including itself, by  $\mathcal{N}(v)$ . We claim that the total charge on the vertices of  $\mathcal{N}(v)$  is at most  $\ln \Delta$ . Initially, none of the vertices in  $\mathcal{N}(v)$  are charged. Let  $u_1, u_2, \dots, u_l$  be the vertices in  $\mathcal{S}$  which charge some vertices of  $\mathcal{N}(v)$  in that order. This charge could either be the direct charge or a transfer of charge from some vertex in  $R$ . For  $i \in [l]$ , let  $O_i \subseteq \mathcal{N}(v)$  denote the set of vertices that remain uncharged (either directly or through a transfer), after the vertex  $u_i$  is picked into  $\mathcal{S}$ . Let  $O_0 = \mathcal{N}(v)$ .

We will now show that, for every  $u_i$ ,  $|C_{u_i}| \geq |O_{i-1}|$ . Let us consider the iteration of the greedy algorithm in which  $u_i$  is picked. We claim that none of the vertices in  $O_{i-1}$  can be dominated by any vertex chosen before  $u_i$  in the greedy algorithm. Let  $w \in O_{i-1}$  be some vertex which is dominated by some vertex  $u'$  chosen by greedy before  $u_i$ , such that  $w \in C_{u'}$ . Clearly  $u' \in L'_1 \cup L'_2 \cup L'_3$  should hold, because no vertex in  $R \setminus L_3$  can dominate  $w$ . But since  $u'$  was chosen before  $u_i$  and  $u' \in L'_1 \cup L'_2 \cup L'_3$ ,  $u'$  must be chosen into  $\mathcal{S}$  before  $u_i$ . Hence,  $w$  cannot be an uncharged vertex in the current iteration leading to a contradiction.

Thus, in the iteration where the greedy algorithm was about to choose  $u_i$ , none of the vertices  $O_{i-1}$  have been dominated. Hence if the greedy were to choose  $v$ , then  $p(v) \geq |O_{i-1}|$ . Since the greedy algorithm chooses vertex  $u_i$  instead of  $v$ , we have  $|C_{u_i}| \geq |O_{i-1}|$ .

The total charge in this iteration ( $C_{u_i} \cap \mathcal{N}(v)$ ) is thus at most  $\frac{|O_{i-1}| - |O_i|}{|O_{i-1}|}$ . Adding these charges over all  $l$  iterations, we get, using an analysis very similar to the set cover analysis [19],  $\sum_{w \in \mathcal{N}(v)} \rho(w) \leq H(\Delta)$ , where  $H$  is the harmonic function and  $\Delta$  is the maximum degree. Adding up the charges over all vertices in  $L_1$ , we get  $\sum_{u \in C_{12} \cup UC_{12}} \rho(u) \leq \sum_{v \in L_1} \sum_{w \in \mathcal{N}(v)} \rho(w) \leq$

$|\text{OPT}| \ln \Delta$ . Hence we have  $|\mathcal{S}| \leq |\text{OPT}| \ln \Delta$ . Since  $\mathcal{S}$  was a maximal set having profit at most  $n'$ , we obtain a set  $D'$  with  $|D'| = |\mathcal{S}| + 1$  with profit at least  $n'$  by adding a single vertex to  $\mathcal{S}$ , which gives us the desired result.

**THEOREM 4.1.** *Let  $\text{OPT}$  be the optimal solution set for an instance of PCDS. There exists a tree  $\hat{T}$  with at most  $2|\text{OPT}| \ln \Delta + |\text{OPT}| + 1$  edges such that  $\sum_{v \in \hat{T}} p(v) \geq n'$ .*

*Proof.* In Lemma 4.1, we have shown that there exists a subset  $D' \subseteq L'_1 \cup L'_2 \cup L'_3$  of size  $|\text{OPT}| \ln \Delta + 1$  that has profit at least  $n'$ . However this set  $D'$  need not be connected. We now show that this set  $D'$  can be connected without paying too much. Firstly we note that for every vertex  $v \in L_3 \cap D'$ , there exists a vertex  $w \in L_2$  such that  $w$  dominates  $v$ . Thus we can pick a subset  $D'' \subseteq L_2$  of size at most  $|L_3 \cap D'| \leq |\text{OPT}| \ln \Delta + 1$  which dominates all vertices of  $L_3 \cap D'$ . Now, it is sufficient to ensure that all the vertices of  $(D' \cap L_2) \cup D''$  are connected. This can be achieved by simply adding all the vertices of  $L_1$  to our solution. Thus we have shown that  $\hat{D} = D' \cup D'' \cup L_1$  induces a connected subgraph with profit at least  $n'$  and the number of vertices in  $\hat{D} \leq |D'| + |D''| + |L_1| \leq 2|\text{OPT}| \ln \Delta + |\text{OPT}| + 2$ . Hence there exists subtree  $\hat{T}$  on these vertices with at most  $(2 \ln \Delta + 1)|\text{OPT}| + 1$  edges with the requisite total profit.

**COROLLARY 4.1.** *Algorithm 4.1 is a  $4 \ln \Delta + 2 + o(1)$ -approximation algorithm for PCDS.*

*Proof.* Let  $\text{OPT}$  be the optimal solution of the PCDS instance. As per Theorem 4.1, we know that there exists a Steiner tree  $\hat{T}$  with at most  $2|\text{OPT}| \ln \Delta + |\text{OPT}| + 1$  edges whose total profit exceeds the quota  $n'$ . Hence, the tree  $T$  returned by the 2-approximation for the QST problem has at most  $4|\text{OPT}| \ln \Delta + 2|\text{OPT}| + 2$  edges. Thus, we obtain a  $4 \ln \Delta + 2 + o(1)$  approximation algorithm.

## 5 Budgeted Connected Dominating Set

We now turn our attention to the Budgeted Connected Dominating Set (BCDS) problem. We recall that in the BCDS problem, we have to choose at most  $k$  vertices that induce a connected subgraph and maximize the number of dominated vertices.

**5.1 Algorithm** Algorithm 5.1 is very similar to the one we used to obtain a partial connected dominating set. We start by running the standard greedy algorithm to find a dominating set  $D$  in the graph. We set the profits of vertices in  $D$  as the number of newly covered vertices at each step of the greedy algorithm, while we

assign zero profit for the remaining vertices in  $V \setminus D$ . In the analysis section, we show that there is a tree on at most  $3k$  vertices that has a total profit of at least  $(1 - \frac{1}{e})\text{OPT}$  where  $\text{OPT}$  is the number of vertices dominated by an optimal solution. Note that we may assume that we have guessed  $\text{OPT}$  by trying out values between  $k$  and  $n$  using, say, binary search. We run the 2 approximation algorithm for the Quota Steiner tree problem on this instance with the quota being set to  $(1 - \frac{1}{e})\text{OPT}$ . This will result in a tree with at most  $6k$  nodes with total profit at least  $(1 - \frac{1}{e})\text{OPT}$ . Thus we obtain a  $(6, 1 - \frac{1}{e})$  bicriteria approximation algorithm. To convert this bicriteria approximation into a true approximation, we use a dynamic program (Section 5.2.2) to find the “best” subtree on at most  $k$  vertices from this tree of  $6k$  vertices. We use a simple tree decomposition scheme to show that the best tree dominates at least  $\frac{1}{13}(1 - \frac{1}{e})\text{OPT}$  nodes.

---

ALGORITHM 5.1. GREEDY PROFIT LABELING ALGORITHM FOR BCDS.

**Input:** Graph  $G = (V, E)$  and  $k \in \mathbb{N}$ .

**Output:** Tree  $\tilde{T}$  with cost at most  $k$ .

---

- 1: Compute the greedy dominating set  $D$  and the corresponding profit function  $p : V \rightarrow \mathbb{N}$  using the Algorithm 4.2.
  - 2:  $\text{OPT} \leftarrow$  number of vertices dominated by an optimal solution.  $\triangleright$  Guess using binary search between  $k$  and  $n$
  - 3: Use the 2-approximation for QST problem [37] to obtain a tree  $T$  with profit at least  $(1 - \frac{1}{e})\text{OPT}$ .  $\triangleright$  We show that  $|T| \leq 6k$ .
  - 4: Use the dynamic program of Section 5.2.2 to find  $\tilde{T}$ , the best subtree of  $T$  having at most  $k$  vertices.
- 

**5.2 Analysis** Let  $L_1$  denote the vertices in an optimal solution. Let layers  $L_2, L_3, R$ , and  $L'_i$  be defined as in Section 4.  $\text{OPT} = |L_1 \cup L_2|$  is the number of vertices dominated by the optimal solution.

Let  $L'_1 \cup L'_2 \cup L'_3 = \{v_1, v_2, \dots, v_l\}$  where the vertices are according to the order in which they were selected by the greedy algorithm. Let  $D' = \{v_1, v_2, \dots, v_k\}$  denote the first  $k$  vertices from  $L'_1 \cup L'_2 \cup L'_3$ . In Lemma 5.1, we prove that the total profit of  $D' = \sum_{v \in D'} p(v)$  is at least  $(1 - \frac{1}{e})\text{OPT}$ . Next, we can show that these  $k$  vertices can be connected by using at most  $2k$  more vertices, thus proving the existence of a tree with at most  $3k$  vertices having the desired total profit.

Let  $g_i$  denote the total profit after picking the first  $i$  vertices from  $D'$ , i.e.,  $g_i = \sum_{j=1}^i p(v_j)$ . We start by proving that the following recurrence holds for every  $i = 0$  to  $k - 1$ .

CLAIM 1.  $g_{i+1} - g_i \geq \frac{1}{k}(\text{OPT} - g_i)$

*Proof.* Consider the iteration of the greedy algorithm, where vertex  $v_{i+1}$  is being picked. We first show that at most  $g_i$  vertices of  $L_1 \cup L_2$  have been already been dominated. Note that any vertex  $w \in L_1 \cup L_2$  that has been already dominated must have been dominated by a vertex in  $\{v_1, v_2, \dots, v_i\}$ . This is because no vertex from  $R \setminus L_3$  can neighbor  $w$ . Since  $g_i = \sum_{j=1}^i p(v_j)$  is the total profit gained so far, it follows that at most  $g_i$  vertices from  $L_1 \cup L_2$  have been dominated. Hence we have that there are at least  $\text{OPT} - g_i$  undominated vertices in  $L_1 \cup L_2$ . Since the  $k$  vertices of  $L_1$  together dominate all of these, it follows that there exists at least one vertex  $v \in L_1$  which neighbors at least  $\frac{1}{k}(\text{OPT} - g_i)$  undominated vertices.

We conclude this proof by noting that since the greedy algorithm chose to pick  $v_{i+1}$  at this stage, instead of the  $v$  above, it follows that  $p(v_{i+1}) = g_{i+1} - g_i \geq \frac{1}{k}(\text{OPT} - g_i)$ .

LEMMA 5.1. *Let  $\text{OPT}$  be the number of vertices dominated by an optimal solution for BCDS. Then there exists a subset  $D' \subseteq D$  of size  $k$  with total profit at least  $(1 - \frac{1}{e})\text{OPT}$ . Further,  $D'$  can be connected using at most  $2k$  Steiner vertices.*

*Proof.* From the Claim 1, the profit after  $i + 1$  iterations is given by

$$g_{i+1} \geq \frac{\text{OPT}}{k} + g_i \left(1 - \frac{1}{k}\right).$$

By solving this recurrence, we get  $g_i \geq (1 - (1 - \frac{1}{k})^i)\text{OPT}$ . Hence, we obtain the following.

$$\sum_{v \in D'} p(v) = g_k \geq (1 - (1 - \frac{1}{k})^k)\text{OPT} \geq (1 - \frac{1}{e})\text{OPT}$$

We show that  $D'$  can be connected by at most  $2k$  Steiner nodes to form a connected tree. Note that for every vertex  $v \in L_3 \cap D'$ , there exists a vertex  $w \in L_2$  such that  $w$  neighbors  $v$ . Thus we can pick a subset  $D'' \subseteq L_2$  of size at most  $|L_3 \cap D'| \leq k$  which dominates all vertices of  $L_3 \cap D'$ . Now, it is sufficient to ensure that all the vertices of  $(D' \cap L_2) \cup D''$  are connected. This can be achieved by simply adding all the  $k$  vertices of  $L_1$ . Thus we have shown that  $\hat{D} = D' \cup D'' \cup L_1$  induces a connected subgraph with profit at least  $(1 - \frac{1}{e})\text{OPT}$  and  $|\hat{D}| \leq |D'| + |D''| + |L_1| \leq 3k$ .

LEMMA 5.2. *There is a  $(6, (1 - \frac{1}{e}))$  bicriteria approximation algorithm for the BCDS problem.*

*Proof.* Lemma 5.1 shows that there exists a Steiner tree with at most  $3k$  vertices having total profit greater

than a quota of  $(1 - \frac{1}{e})\text{OPT}$ . Hence, using the 2-approximation for the QST problem, we obtain a tree  $T$  of at most  $6k$  nodes and total profit at least  $(1 - \frac{1}{e})\text{OPT}$ . Thus we obtain a  $(6, (1 - \frac{1}{e}))$  bicriteria approximation algorithm for the BCDS problem.

**5.2.1 Converting the Bicriteria Approximation to a True Approximation** In order to obtain a true approximate solution (solution of size  $k$ ), we need a technique to find a small subtree  $\tilde{T} \subseteq T$  of  $k$  vertices which has high total profit. In Section 5.2.2, we show that this problem can be easily solved in polynomial time using dynamic programming. However, simply finding the subtree which maximizes the profit is not enough to give a good approximation ratio. We need a way to compare the total profit of the subtree  $\tilde{T}$  with the entire profit  $P = \sum_{v \in T} p(v)$ . We now show that if  $n = 6k$ , we can obtain a subtree having profit at least  $\frac{1}{13}P$ .

The following lemma is well known in folklore and can be easily proven by induction. It can also be seen as an easy consequence of a theorem by Jordan [38].

**LEMMA 5.3.** (JORDAN [38]) *Given any tree on  $n$  vertices, we can decompose it into two trees (by replicating a single vertex) such that the smaller tree has at most  $\lceil \frac{n}{2} \rceil$  nodes and the larger tree has at most  $\lceil \frac{2n}{3} \rceil$  nodes.*

We now show the following -

**LEMMA 5.4.** *Let  $k$  be greater than a sufficiently large constant. Given a tree  $T$  with  $6k$  nodes, we can decompose it into 13 trees of size at most  $k$  nodes each.*

*Proof.* We use Lemma 5.3 to decompose the tree into two trees  $T_1$  and  $T_2$  such that  $|T_1| \leq |T_2|$ . In this decomposition, at most one vertex is duplicated, therefore  $|T_1| + |T_2| \leq 6k + 1$ . Also, we have  $|T_1| \leq 3k$ . We now have two cases:

**Case 1:**  $|T_1| \geq 3k - 1$ . In this case,  $|T_2| \leq 6k + 1 - |T_1| \leq 3k + 2$ . Now repeatedly using Lemma 5.3 we can see that  $T_1$  can be decomposed into at most 6 trees and  $T_2$  can be decomposed into at most 7 trees of size at most  $k$ . This is shown in the Figure 3. Hence, in this case, we can decompose the tree  $T$  into 13 trees.

**Case 2:**  $|T_1| \leq 3k - 2$ . In this case,  $|T_2| \leq 4k$ . In this case, we can decompose  $T_1$  into 5 trees and  $T_2$  can be decomposed into 8 trees. This is shown in Figure 4. Thus in this case, we can decompose  $T$  into 13 trees.

Using Lemma 5.4, we can convert the bicriteria approximation for BCDS to a true approximation algorithm. In particular, we show the following -

**THEOREM 5.1.** *There is a  $\frac{1}{13}(1 - \frac{1}{e})$  approximation algorithm for the BCDS problem.*

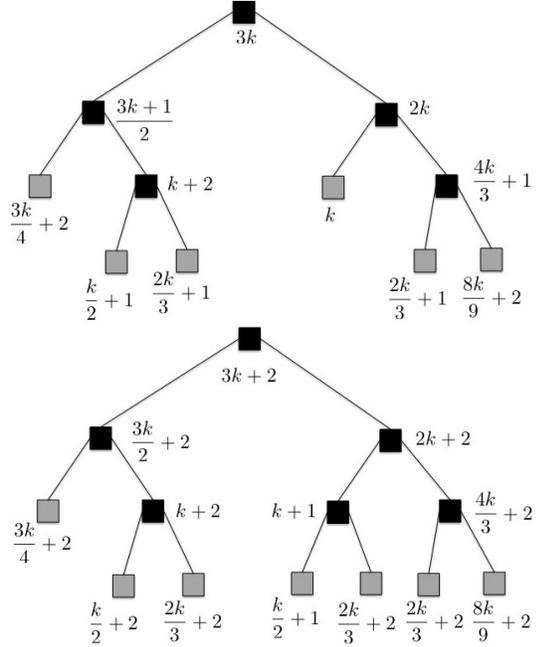


Figure 3: First tree decomposes into 6 subtrees and second tree decomposes into 7 trees. In total, we obtain 13 subtrees. The number associated with each node is the upper bound on the size of the subtree.

*Proof.* By Lemma 5.2, we obtain a tree  $T$  with at most  $6k$  nodes with profit  $(1 - \frac{1}{e})\text{OPT}$ . Now using Lemma 5.4, we obtain 13 trees in the worst case, say  $T_1, T_2, \dots, T_{13}$ . Finally, out of these 13 trees (each of size at most  $k$ ), we pick the tree  $\tilde{T}$  with the highest total profit. Let,  $p(T) = \sum_{v \in T} p(v)$  denote the total profit of tree  $T$ . Then we have,

$$p(\tilde{T}) \geq \frac{1}{13} \sum_{i=1}^{13} p(T_i) \geq \frac{1}{13} p(T) \geq \frac{1}{13} (1 - \frac{1}{e}) \text{OPT}$$

Thus we have a  $\frac{1}{13}(1 - \frac{1}{e})$  approximation guarantee.

**5.2.2 Finding the Best Subtree** Although the decomposition Lemma 5.4 is useful to prove a theoretical bound, from a practical perspective it is better to use a dynamic programming approach to find the best  $k$  sub-tree. Formally, we have the following problem. Given a tree  $T = (V, E)$  of  $n$  vertices, profits on vertices  $p : V \rightarrow \mathbb{Z}^+ \cup \{0\}$ , and an integer  $k$ , find a subtree  $\tilde{T}$  of  $k$  vertices which maximizes the total profit  $\tilde{P} = \sum_{v \in \tilde{T}} p(v)$ . We show that this problem can be solved in polynomial time using dynamic programming. Let the tree  $T$  be rooted at an arbitrary vertex and  $T_v$  denote the subtree rooted at a vertex  $v$ . We define the following -

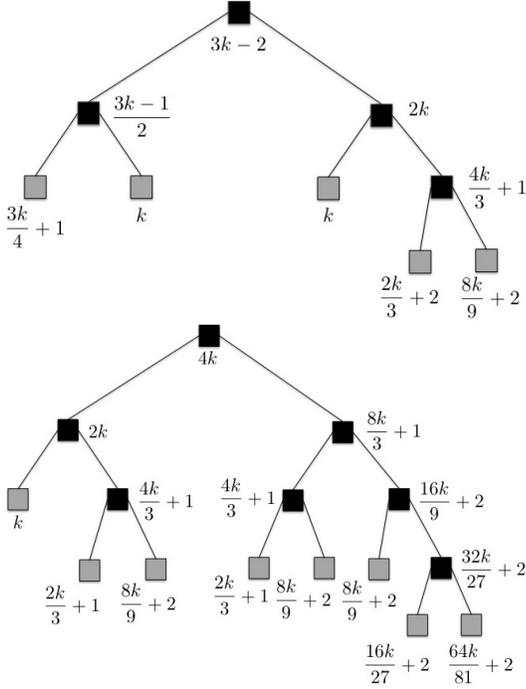


Figure 4: First tree decomposes into 5 subtrees and second tree decomposes into 8 trees. In total, we obtain 13 subtrees. The number associated with each node is the upper bound on the size of the subtree.

$F(v, i) \leftarrow$  best solution of at most  $i$  vertices completely contained inside  $T_v$ .

$G(v, i) \leftarrow$  best solution of at most  $i$  vertices completely contained inside  $T_v$  such that  $v$  is a part of the solution.

The desired solution is thus at  $F(\text{root}, k)$ . The base cases (when  $v$  is a leaf) are trivial. Let  $v_1, v_2, \dots, v_l$  denote the children of vertex  $v$ . We now have the following recurrence -

$$F(v, i) = \max \left\{ \max_{1 \leq j \leq l} \{F(v_j, i)\}, G(v, i) \right\}$$

$$G(v, i) = p(v) + M(l, i - 1)$$

where  $M(j, i')$  denotes the best way to distribute a budget of  $i'$  among the first  $j$  children of  $v$ . In other words,

$$M(l, i - 1) = \max_{i_1 + i_2 + \dots + i_l = i - 1} \left\{ \sum_j G(v_j, i_j) \right\}$$

$M(j, i')$  is computed using another dynamic program as follows. Again the base cases when  $j = 0$  or  $i' = 0$  are trivial. For  $1 \leq j \leq l$  and  $1 \leq i' \leq i - 1$ , we have the

following recurrence -

$$M(j, i') = \max_{0 \leq i^* \leq i'} \{M(j - 1, i^*) + G(v_j, i' - i^*)\}$$

## 6 Budgeted Generalized CDS

In this section, we show that our approach extends to more general budgeted connected domination problems. Formally, given a graph  $G = (V, E)$ , a budget  $k$ , and a monotone *special submodular* profit function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup 0$ , find a subset  $S \subseteq V$  which maximizes  $f(S)$  such that  $|S| \leq k$  and induces a connected subgraph of  $G$ . As mentioned earlier in Section 2, this problem captures the budgeted variants of the capacitated and weighted profit connected dominating set problems.

**6.1 Algorithm.** Algorithm 6.1 begins by running the standard greedy algorithm to find a basis of the polymatroid associated with  $f$ . In other words, we greedily pick a vertex  $v$  with the maximum marginal profit  $f(D \cup \{v\}) - f(D)$  until all vertices have zero marginal profit. With every selected vertex, we associate the marginal profit gained, and associate zero profit with the other vertices. Finally, we run a quota Steiner tree algorithm using these profits to find the smallest tree that yields a profit of at least  $(1 - \frac{1}{e})\text{OPT}$  where  $\text{OPT}$  is the optimal profit (which we guess). In the analysis section, we show that there exists a tree  $\hat{T}$  of size at most  $3k$  with  $f(\hat{T}) \geq (1 - \frac{1}{e})\text{OPT}$ . Hence, the 2-approximation for the quota Steiner tree yields a tree  $T$  of size at most  $6k$  yielding the desired profit. Finally using the tree decomposition described earlier, we show that we can obtain a tree  $\tilde{T}$  of size at most  $k$  with  $f(\tilde{T}) \geq \frac{1}{13}(1 - \frac{1}{e})\text{OPT}$ .

**6.2 Analysis.** Let the  $L_1$  denote the vertices in the optimal solution and  $f(L_1) = \text{OPT}$ . Let  $L_2$  denote the set of vertices which have at least one neighbor in  $L_1$ , and similarly let  $L_3$  denote the set of vertices having a neighbor in  $L_2$  (and NOT in  $L_1$ ). Let  $R = V \setminus \{L_1 \cup L_2 \cup L_3\}$  denote the rest of the vertices. Let  $L'_i = D \cap L_i$  where  $D$  is the set of vertices chosen by the greedy algorithm.

Further, let  $D'$  denote the first  $k$  vertices picked by the greedy algorithm from  $L'_1 \cup L'_2 \cup L'_3$ . To simplify notation, let  $D' = \{v_1, v_2, \dots, v_k\}$  and let  $D_i$  denote the set of vertices already picked by the greedy algorithm when the vertex  $v_{i+1}$  is being chosen. Hence we have  $v_{i+1} = \arg \max_{v \in V \setminus D_i} f(D_i \cup \{v\}) - f(D_i)$  and  $p(v_{i+1}) = f(D_i \cup \{v_{i+1}\}) - f(D_i)$ . Note that in particular  $D_i \subseteq D$  but may not be a subset of  $D'$ . Also let  $D'_i = \cup_{j=1}^i v_j$  denote the first  $i$  vertices in  $D'$ . Let  $P(D'_i) = \sum_{v \in D'_i} p(v)$  denote the total profit associated with the set  $D'_i$ . Finally let  $D''_i = D_i \setminus D'_i$  be the vertices

in  $D_i \cap R$ .

CLAIM 2.  $p(v_{i+1}) = P(D'_{i+1}) - P(D'_i) \geq \frac{1}{k}(\text{OPT} - P(D'_i))$

*Proof.* Consider the marginal profit of the set  $L_1 \setminus D'_i$ . Since  $N(D''_i)$  does not intersect with  $N(L_1)$ , we have,

$$\begin{aligned}
 f_{D'_i}(L_1 \setminus D'_i) &= f_{D'_i \cup D''_i}(L_1 \setminus D'_i) \\
 &= f_{D'_i}(D'_i \cup (L_1 \setminus D'_i)) - f_{D''_i}(D'_i) \\
 &\geq f_{D''_i}(L_1) - f_{D''_i}(D'_i) \\
 &= f(L_1) - f_{D''_i}(D'_i) \\
 (6.1) \quad &= \text{OPT} - f_{D''_i}(D'_i)
 \end{aligned}$$

Let us now consider the term  $f_{D''_i}(D'_i)$ . Adding up over successive marginal profits,

$$\begin{aligned}
 (6.2) \quad f_{D''_i}(D'_i) &= \sum_{j=1}^i f_{D''_i \cup D'_{j-1}}(v_j) \leq \sum_{j=1}^i f_{D_{j-1}}(v_j) \\
 &= \sum_{j=1}^i p(v_j) = P(D'_i)
 \end{aligned}$$

From Eq (6.1) and Eq (6.2),

$$f_{D'_i}(L_1 \setminus D'_i) \geq \text{OPT} - P(D'_i)$$

As  $f$  is submodular, we have

$$f_{D'_i}(L_1 \setminus D'_i) \leq \sum_{w \in L_1 \setminus D'_i} f_{D'_i}(\{w\})$$

Since  $|L_1 \setminus D'_i| \leq k$ , there exists at least one vertex  $w \in L_1 \setminus D'_i$  satisfying

$$f_{D'_i}(\{w\}) \geq \frac{1}{k} f_{D'_i}(L_1 \setminus D'_i) \geq \frac{1}{k}(\text{OPT} - P(D'_i))$$

Using  $f_{D_i}(\{w\}) = f_{D'_i}(\{w\})$  and the fact that greedy picked  $v_{i+1}$  at this stage

$$\begin{aligned}
 p(v_{i+1}) &= f_{D_i}(\{v_{i+1}\}) \geq f_{D_i}(\{w\}) \\
 &\geq \frac{1}{k}(\text{OPT} - P(D'_i))
 \end{aligned}$$

Solving the recurrence of Claim 2, we have  $P(D') \geq (1 - \frac{1}{e})\text{OPT}$ .

We thus have a set  $D'$  of size  $k$  which yields a total profit of at least  $(1 - \frac{1}{e})\text{OPT}$ . We now proceed to show that the above set  $D'$  can be connected at a relatively low cost. Since every vertex in  $D'$  can be connected to  $L_1$  using at most one vertex (from  $L_2$ ), we can obtain a connected subset  $\hat{T}$  of size at most  $3k$  by choosing  $D'$ ,  $L_1$  and vertices in  $L_2$  as described. Hence, the 2-approximation for the QST problem will yield a tree  $T$  of size at most  $6k$  which would give a profit of at least  $(1 - \frac{1}{e})\text{OPT}$ . Finally applying the tree decomposition described earlier we obtain a tree  $\tilde{T}$  of size  $\leq k$  with  $f(\tilde{T}) \geq P(\tilde{T}) \geq \frac{1}{13}(1 - \frac{1}{e})\text{OPT}$ .

---



---

ALGORITHM 6.1. GREEDY PROFIT LABELING ALGORITHM FOR BGCDS.

**Input:** Graph  $G = (V, E)$ , a monotone special submodular function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup \{0\}$  and  $k \in \mathbb{Z}^+ \cup \{0\}$ .

**Output:** Tree  $\tilde{T}$  with at most  $k$  vertices.

---

- 1: Run the Generalized Greedy Dominating Set Routine (Algorithm 6.2) on  $(G, f)$  to obtain a subset  $D$  and a profit function  $p : V \rightarrow \mathbb{N}$ .
  - 2:  $\text{OPT} \leftarrow$  profit of an optimal solution. (Guess using binary search 0 and  $f(V)$ ).
  - 3:  $T \leftarrow$  2-approximation for QST with quota  $(1 - \frac{1}{e})\text{OPT}$ .
  - 4: Use the dynamic program of Section 5.2.2 to find  $\tilde{T}$ , the best subtree of  $T$  having at most  $k$  vertices.
- 
- 

---



---

ALGORITHM 6.2. GENERALIZED GREEDY DOMINATING SET.

**Input:** Graph  $G = (V, E)$  and a monotone special submodular function  $f : 2^V \rightarrow \mathbb{Z}^+ \cup \{0\}$ .

**Output:**  $D \subseteq V$  such that  $f(D) = f(V)$  and profit function  $p : V \rightarrow \mathbb{Z}^+ \cup \{0\}$ .

---

- 1:  $D \leftarrow \phi$
  - 2: **while**  $f(D) \neq f(V)$  **do**
  - 3:      $v \leftarrow \arg \max_{v \in V \setminus D} f(D \cup \{v\}) - f(D)$
  - 4:      $p(v) \leftarrow f(D \cup \{v\}) - f(D)$
  - 5:      $D \leftarrow D \cup \{v\}$
  - 6: **end while**
  - 7: **for all**  $v \in V \setminus D$  **do**
  - 8:      $p(v) \leftarrow 0$
  - 9: **end for**
- 
- 

## 7 Partial Generalized Connected Domination

We now consider a partial coverage version of the generalized connected domination presented in Section 6. In this problem, the goal is to find the smallest subset of vertices which induce a connected subgraph and have total profit at least  $q$  (quota). Just as for the budgeted case, the algorithm proceeds by finding a spanning subset greedily. Using profits as defined by the greedy algorithm, we then find a QST having total profit at least  $q$ . In the analysis section, we show that there exists a tree  $\hat{T}$  of size at most  $2k \ln q + k$  with total profit at least  $q$ . Hence, the 2-approximation for QST yields a tree  $T$  of size at most  $4k \ln q + 2k$  leading to a  $O(4 \ln q)$  approximation.

**7.1 Analysis** We reuse notation from Section 6 regarding the layers  $L_i$  and  $L'_i$ . Let  $D'$  denote the first

$k \ln q + 1$  vertices picked by the greedy algorithm from  $L'_1 \cup L'_2 \cup L'_3$ . We now show that the total profit of vertices in  $D'$  is at least  $q$ .

CLAIM 3.  $P(D') \geq q$

*Proof.* As per Claim 2, we obtain the following recurrence

$$(7.3) \quad P(D'_{i+1}) \geq (1 - (1 - \frac{1}{k})^{i+1})q$$

Substituting  $i + 1 = k \ln q$ , we get,

$$(7.4) \quad P(D'_{k \ln q}) \geq (1 - (1 - \frac{1}{k})^{k \ln q})q$$

$$(7.5) \quad \geq (1 - \frac{1}{q})q \geq q - 1$$

Since profit function  $f$  is integral, we have

$$(7.6) \quad P(D'_{k \ln q + 1}) \geq q$$

THEOREM 7.1. *Given that the optimal solution is of size  $k$ , there exists a tree  $\hat{T}$  of size at most  $k \ln q + k + 2$  such that  $\sum_{v \in \hat{T}} p(v) \geq q$*

*Proof.* In Claim 3 above, we have demonstrated the existence of a set of size at most  $k \ln q + 1$  with the requisite total profit. We now show that this set can be connected at low cost. As in Theorem 4.1, we can see that by selecting at most  $k \ln q + 1$  more vertices from layer  $L_2$  and at most  $k$  vertices from layer  $L_1$ , the set  $D'$  can be connected to form a tree  $\hat{T}$ .

Finally using the 2-approximation for QST, we obtain a  $O(4 \ln q)$  approximation.

## 8 Conclusion and Future Work

We consider partial and budgeted versions of the well studied connected dominating set problem. We observe that various algorithms which perform well in the *complete* version of the connected dominating set have unbounded approximation guarantee in the partial case. Using a surprising *greedy profit labeling* algorithm we obtain the first  $O(\log n)$  approximation for the partial connected dominating set problem and a  $\frac{1}{13}(1 - \frac{1}{e})$  approximation for the budgeted version. We also extend our results to a *special submodular* problem, which includes capacitated and weighted profit versions of the PCDS and BCDS problems as special cases. Our results are tight up to a constant factor in all the cases. A natural open question is to improve these constants.

**Acknowledgment:** The first author would like

to thank Yossi Azar for useful discussions, held during the Dagstuhl seminar on Scheduling (2013), about the failure of prior methods for the budgeted CDS problem.

## References

- [1] Sanjeev Arora and George Karakostas. A  $2 + \epsilon$  approximation algorithm for the  $k$ -MST problem. In *SODA*, pages 754–759, 2000.
- [2] Konstantin Avrachenkov, Prithwish Basu, Giovanni Neglia, Bruno F. Ribeiro, and Don Towsley. Online Myopic Network Covering. *CoRR*, abs/1212.5035, 2012.
- [3] Baruch Awerbuch, Yossi Azar, Avrim Blum, and Santosh Vempala. Improved approximation guarantees for minimum-weight  $k$ -trees and prize-collecting salesmen. In *STOC*, pages 277–283, 1995.
- [4] Reuven Bar-Yehuda. Using homogenous weights for approximating the partial cover problem. In *SODA*, pages 71–75, 1999.
- [5] Reuven Bar-Yehuda, Guy Flysher, Julián Mestre, and Dror Rawitz. Approximation of partial capacitated vertex cover. In *ESA*, pages 335–346, 2007.
- [6] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *FOCS*, pages 184–193, 1996.
- [7] MohammadHossein Bateni, MohammadTaghi Hajiaghayi, and Vahid Liaghat. Improved Approximation Algorithms for (Budgeted) Node-weighted Steiner Problems. In *ICALP*, pages 81–92, 2013.
- [8] Avrim Blum, R. Ravi, and Santosh Vempala. A constant-factor approximation algorithm for the  $k$  MST problem (extended abstract). In *STOC*, pages 442–448, 1996.
- [9] Christian Borgs, Michael Brautbar, Jennifer Chayes, Sanjeev Khanna, and Brendan Lucier. The power of local information in social networks. In *Internet and Network Economics*, pages 406–419. Springer, 2012.
- [10] Nader H Bshouty and Lynn Burroughs. Massaging a linear programming solution to give a 2-approximation for a generalization of the vertex cover problem. In *STACS*, pages 298–308, 1998.
- [11] G. Calinescu and A. Zelikovsky. The polymatroid Steiner problems. *Journal of Combinatorial Optimization*, 9:281–294, 2005.
- [12] Moses Charikar and Samir Khuller. A robust maximum completion time measure for scheduling. In *SODA*, pages 324–333, 2006.
- [13] Moses Charikar, Samir Khuller, David M. Mount, and Giri Narasimhan. Algorithms for facility location problems with outliers. In *SODA*, pages 642–651, 2001.
- [14] Chandra Chekuri, Guy Even, and Guy Kortsarz. A greedy approximation algorithm for the group Steiner problem. *Discrete Applied Mathematics*, 154(1):15–34, 2006.
- [15] Ke Chen. A constant factor approximation algorithm for  $k$ -median clustering with outliers. In *SODA*, pages 826–835, 2008.

- [16] Maggie X Cheng, Lu Ruan, and Weili Wu. Achieving minimum coverage breach under bandwidth constraints in wireless sensor networks. In *INFOCOM*, pages 2638–2645, 2005.
- [17] Maggie X Cheng, Lu Ruan, and Weili Wu. Coverage breach problems in bandwidth-constrained sensor networks. *TOSN*, page 12, 2007.
- [18] Xiuzhen Cheng, Xiao Huang, Deying Li, Weili Wu, and Ding-Zhu Du. A polynomial-time approximation scheme for the minimum-connected dominating set in ad hoc wireless networks. *Networks*, 42(4):202–208, 2003.
- [19] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to Algorithms, 3<sup>rd</sup> Edition. page 977, 2009.
- [20] Bevan Das and Vaduvur Bharghavan. Routing in ad-hoc networks using minimum connected dominating sets. In *IEEE International Conference on Communications (ICC)*, volume 1, pages 376–380. IEEE, 1997.
- [21] Erik D Demaine and MohammadTaghi Hajiaghayi. Bidimensionality: New Connections between FPT algorithms and PTASs. In *SODA*, pages 590–601, 2005.
- [22] D.Z. Du and P.J. Wan. *Connected Dominating Set: Theory and Applications*. Springer Optimization and Its Applications. Springer New York, 2013.
- [23] Devdatt P. Dubhashi, Alessandro Mei, Alessandro Panconesi, Jaikumar Radhakrishnan, and Aravind Srinivasan. Fast distributed algorithms for (weakly) connected dominating sets and linear-size skeletons. In *SODA*, pages 717–724, 2003.
- [24] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC*, pages 448–455, 2003.
- [25] Uriel Feige. A threshold of  $\ln n$  for approximating set cover. *J. ACM*, 45(4):634–652, July 1998.
- [26] Rajiv Gandhi, Samir Khuller, and Aravind Srinivasan. Approximation algorithms for partial covering problems. *J. Algorithms*, 53(1):55–84, 2004.
- [27] Rajiv Gandhi and Srinivasan Parthasarathy. Distributed algorithms for connected domination in wireless networks. *Journal of Parallel and Distributed Computing*, 67(7):848–862, 2007.
- [28] Naveen Garg. A 3-approximation for the minimum tree spanning  $k$  vertices. In *FOCS*, pages 302–309, 1996.
- [29] Naveen Garg. Saving an epsilon: a 2-approximation for the  $k$ -MST problem in graphs. In *STOC*, pages 396–402, 2005.
- [30] Naveen Garg, Goran Konjevod, and R. Ravi. A Polylogarithmic Approximation Algorithm for the Group Steiner Tree Problem. In *SODA*, pages 253–259, 1998.
- [31] Teofilo F. Gonzalez. Clustering to Minimize the Maximum Intercluster Distance. *Theor. Comput. Sci.*, 38:293–306, 1985.
- [32] Sudipto Guha and Samir Khuller. Approximation algorithms for connected dominating sets. *Algorithmica*, 20(4):374–387, 1998.
- [33] Anupam Gupta, Ravishankar Krishnaswamy, Amit Kumar, and Danny Segev. Scheduling with Outliers. In *APPROX-RANDOM*, pages 149–162, 2009.
- [34] Eran Halperin and Robert Krauthgamer. Polylogarithmic inapproximability. In *STOC*, pages 585–594, 2003.
- [35] Eran Halperin and Aravind Srinivasan. Improved approximation algorithms for the partial vertex cover problem. In *Approximation Algorithms for Combinatorial Optimization*, pages 161–174. 2002.
- [36] Dorit S Hochbaum and David B Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM (JACM)*, 33(3):533–550, 1986.
- [37] David S Johnson, Maria Minkoff, and Steven Phillips. The prize collecting Steiner tree problem: theory and practice. In *SODA*, pages 760–769, 2000.
- [38] Camille Jordan. Sur les assemblages de lignes. *Journal für die reine und angewandte Mathematik*, 70:185–190, 1869.
- [39] Michael J Kearns. *The computational complexity of machine learning*. The MIT Press, 1990.
- [40] Samir Khuller, Anna Moss, and Joseph (Seffi) Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39 – 45, 1999.
- [41] Samir Khuller, Barna Saha, and Kanthi K Sarpatwar. New Approximation Results for Resource Replication Problems. In *APPROX-RANDOM*, pages 218–230. 2012.
- [42] Yuzhen Liu and Weifa Liang. Approximate coverage in wireless sensor networks. In *ICN*, pages 68–75, 2005.
- [43] Julián Mestre. A primal-dual approximation algorithm for partial vertex cover: Making educated guesses. *Algorithmica*, 55(1):227–239, 2009.
- [44] Anna Moss and Yuval Rabani. Approximation algorithms for constrained for constrained node weighted Steiner tree problems. In *STOC*, pages 373–382, 2001.
- [45] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions I. *Mathematical Programming*, 14(1):265–294, 1978.
- [46] R. Ravi, R. Sundaram, M. V. Marathe, D. J. Rosenkrantz, and S. S. Ravi. Spanning Trees—Short or Small. *SIAM J. Discret. Math.*, 9(2):178–200, May 1996.
- [47] Petr Slavík. Improved performance of the greedy algorithm for partial cover. *Information Processing Letters*, 64(5):251–254, 1997.
- [48] Aravind Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *FOCS*, pages 588–597, 2001.
- [49] Peng-Jun Wan, Khaled M Alzoubi, and Ophir Frieder. Distributed construction of connected dominating set in wireless ad hoc networks. In *INFOCOM*, volume 3, pages 1597–1604, 2002.
- [50] Laurence A. Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4):385–393, 1982.
- [51] Jie Wu and Hailan Li. On calculating connected dominating set for efficient routing in ad hoc wireless networks. In *DIALM*, pages 7–14, 1999.