Constant factor Approximation Algorithm for Uniform Hard Capacitated Knapsack Median Problem

Sapna Grover
Department of Computer Science, University of Delhi, Delhi, India
sgrover@cs.du.ac.in

Neelima Gupta
Department of Computer Science, University of Delhi, Delhi, India
ngupta@cs.du.ac.in

Samir Khuller
Department of Computer Science, University of Maryland, College Park, MD 20742, USA
samir@cs.umd.edu

Aditya Pancholi
Department of Computer Science, University of Delhi, Delhi, India
apancholi@cs.du.ac.in

Abstract

In this paper, we give the first constant factor approximation algorithm for capacitated knapsack median problem (CKnM) for hard uniform capacities, violating the budget by a factor of $1+\epsilon$ and capacities by a $2+\epsilon$ factor. To the best of our knowledge, no constant factor approximation is known for the problem even with capacity/budget/both violations. Even for the uncapacitated variant of the problem, the natural LP is known to have an unbounded integrality gap even after adding the covering inequalities to strengthen the LP. Our techniques for CKnM provide two types of results for the capacitated k-facility location problem. We present an $O(1/\epsilon^2)$ factor approximation for the problem, violating capacities by $(2+\epsilon)$. Another result is an $O(1/\epsilon)$ factor approximation, violating the capacities by a factor of at most $(1+\epsilon)$ using at most $2k$ facilities for a fixed $\epsilon > 0$. As a by-product, a constant factor approximation algorithm for capacitated facility location problem with uniform capacities is presented, violating the capacities by $(1+\epsilon)$ factor. Though constant factor results are known for the problem without violating the capacities, the result is interesting as it is obtained by rounding the solution to the natural LP, which is known to have an unbounded integrality gap without violating the capacities. Thus, we achieve the best possible from the natural LP for the problem. The result shows that the natural LP is not too bad.

2012 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Capacitated Knapsack Median, Capacitated k -Facility Location

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2018.23

1 Introduction

Facility location and k-median problems are well studied in the literature. In this paper, we study some of their generalizations. In particular, we study capacitated variants of the knapsack median problem (KnM) and the k facility location problem (kFLP). Knapsack median problem is a generalization of the k-median problem, in which we are given a set $\mathcal{C}$ of clients with demands, a set $\mathcal{F}$ of facility locations and a budget $\mathcal{B}$. Setting up a facility

38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2018).
Editors: Sumit Ganguly and Paritosh Pandya; Article No. 23; pp. 23:1–23:21
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

at location i incurs cost f_i (called the *facility opening cost* or simply the *facility cost*) and servicing a client j by a facility i incurs cost $c(i, j)$ (called the *service cost*). We assume that the costs are metric i.e., they satisfy the triangle inequality. The goal is to select the locations to install facilities, so that the total cost for setting up the facilities does not exceed $\mathcal{B}$ and the cost of servicing all the clients by the opened facilities is minimized. When $f_i = 1\ \forall i \in \mathcal{F}$ and $\mathcal{B} = k$, it reduces to the k-median problem. In the *capacitated* version of the problem, we are also given a bound u_i on the maximum number of clients that facility i can serve. Given a set of open facilities, an assignment problem is solved to determine the best way of servicing the clients. Thus any solution is completely determined by the set of open facilities. In this paper, we address the capacitated knapsack median (CKnM) problem with uniform capacities i.e., $u_i = u\ \forall i \in \mathcal{F}$ and clients with unit demands. In particular, we present the following result:

▶ **Theorem 1.** *There is a polynomial time algorithm that approximates hard uniform capacitated knapsack median problem within a constant factor violating the capacity by a factor of at most $(2 + \epsilon)$ and budget by a factor of at most $(1 + \epsilon)$, for every fixed $\epsilon > 0$.*

Our result is nearly the best achievable from rounding the natural LP: we cannot expect to get rid of the violation in the budget as it would imply a constant factor integrality gap for the uncapacitated case which is known to have an unbounded integrality gap. Even with budget violation, capacity violation cannot be reduced to below 2 as it would imply less than 2 factor capacity violation for k-median problem with $k + 1$ facilities. The natural LP has an unbounded integrality gap for this scenario as well[1] [2].

The k-facility location problem (kFLP) is a common generalization of the facility location problem and the k-median problem. In kFLP, we are given a bound k on the maximum number of facilities that can be opened (instead of a budget on the total facility opening cost) and the objective is to minimize the total of facility opening cost and the cost of servicing the clients by the opened facilities. In particular we present the following two results:

▶ **Theorem 2.** *There is a polynomial time algorithm that approximates hard uniform capacitated k-facility location problem within a constant factor $(O(1/\epsilon^2))$ violating the capacities by a factor of at most $(2 + \epsilon)$ for every fixed $\epsilon > 0$.*

▶ **Theorem 3.** *There is a polynomial time algorithm that approximates hard uniform capacitated k-facility location problem within a constant factor $(O(1/\epsilon))$ violating the capacity by a factor of at most $(1 + \epsilon)$ using at most $2k$ facilities for every fixed $\epsilon > 0$.*

As a particular case of CkFLP, we obtain the following interesting result for the capacitated facility location problem (CFLP):

▶ **Corollary 4.** *There is a polynomial time algorithm that approximates hard uniform capacitated facility location problem within a constant factor $(O(1/\epsilon))$ violating the capacity by a factor of at most $(1 + \epsilon)$ for every fixed $\epsilon > 0$.*

---

[1] Let M be a large integer, $u_i = M$ and $k = 2M - 2$. There are M groups of locations; distance between locations within a group is 0 and distance between locations in two different groups is 1. Each group has $2M - 2$ facilities and $2M - 2$ clients, all co-located. In an optimal LP solution each facility is opened to an extent of $1/M$ thereby creating a capacity of $2M - 2$ within each group. In an integer solution, if at most $k + 1 = 2M - 1$ facilities are allowed to be opened then there is at least one group with only one facility opened in it. Thus capacity in the group is M whereas the demand is $2M - 2$. Thus the blowup in capacity is $(2M - 2)/M$.

[2] We thank Moses Charikar for providing the above example where violation in one of the parameters is less than 2 factor and no violation in the other. The example was subsequently modified by us to allow $k + 1$ facilities.

The standard LP is known to have an unbounded integrality gap for CFLP even with uniform capacities. Though constant factor results are known for the problem without violating the capacities [2, 4], our result is interesting as it is obtained by rounding the solution to the natural LP. Our result shows that the natural LP is not too bad.

1.1 Motivation and Challenges

The natural LP for KnM is known to have an unbounded integrality gap [10] even for the uncapacitated case. Obtaining a constant factor approximation for the (capacitated) k-median (CkM) problem is still open, let alone the CKnM problem. Existing solutions giving constant-factor approximation for CkM violate at least one of the two (*cardinality* and *capacity*) constraints. Natural LP is known to have an unbounded integrality gap when any one of the two constraints is allowed to be violated by a factor of less than 2 without violating the other.

Several results [9, 11, 6, 21, 16, 1] have been obtained for CkM that violate either the capacities or the cardinality by a factor of 2 or more. The techniques used for CkM cannot be used for CKnM as they work by transferring the opening from one facility to another (ensuring bounded service cost) facility thereby maintaining the cardinality within claimed bounds. This works well when there are no facility opening costs or the (facility opening) costs are uniform. For the general opening costs, this is a challenge as a facility, good for bounded service cost, may lead to budget violation. To the best of our knowledge, capacitated knapsack median problem has not been addressed earlier.

CkFLP is NP-hard even when there is only one client and there are no facility costs [1]. The hardness results for CkM hold for CkFLP as well. On the other hand, standard LP for capacitated facility location problem (CFLP) has an unbounded integrality gap, thereby implying that constant integrality ratio can not be obtained for CkFLP without violating the capacities even if $k = n$. Byrka *et al.* [6] gave an $O(1/\epsilon^2)$ algorithm for CkFLP when the capacities are uniform (UCkFLP) violating the capacities by a factor of $2 + \epsilon$. They use randomized rounding to bound the expected cost. It can be shown that deterministic pipage rounding cannot be used here. The strength of our techniques is demonstrated in obtaining the first deterministic constant factor approximation with the same capacity violation. The primary source of inspiration for our result in Theorem 3 comes from its corollary.

1.2 Related Work

Capacitated k-median problem has been studied extensively in the literature. For the case of uniform capacities, several results [6, 9, 11, 21, 16] have been obtained that violate either the capacities or the cardinality by a factor of 2 or more. In case of non-uniform capacities, a $(7 + \epsilon)$ algorithm was given by Aardal *et al.* [1] violating the cardinality constraint by a factor of 2 as a special case of Capacitated k-FLP when the facility costs are all zero. Byrka *et al.* [6] gave an $O(1/\epsilon)$ approximation result violating capacities by a factor of $(3 + \epsilon)$.

Li [22] broke the barrier of 2 in cardinality and gave an $\exp(O(1/\epsilon^2))$ approximation using at most $(1 + \epsilon)k$ facilities for uniform capacities. Li gave a sophisticated algorithm using a novel linear program which he calls the *rectangle LP*. The result was extended to non-uniform capacities by the same author using a new LP called *configuration LP* [23]. The approximation ratio was also improved from $\exp(O(1/\epsilon^2))$ to $(O(1/\epsilon^2 \log(1/\epsilon)))$. Though the algorithm violates the cardinality only by $1 + \epsilon$, it introduces a softness bounded by a factor of 2. The running time of the algorithm is $n^{O(1/\epsilon)}$.

Byrka *et al.* [8] broke the barrier of 2 in capacities and gave an $O(1/\epsilon^2)$ approximation violating capacities by a factor of $(1+\epsilon)$ factor for uniform capacities. The algorithm uses randomized rounding to round a fractional solution to the configuration LP. For non-uniform capacities, a similar result has been obtained by Demirci *et al.* [14]. The paper presents an $O(1/\epsilon^5)$ approximation algorithm with capacity violation by a factor of at most $(1+\epsilon)$. The running time of the algorithm is $n^{O(1/\epsilon)}$.

Another closely related problem to Capacitated k-median problem is the Capacitated k-center problem, where-in we have to minimize the maximum distance of a client to a facility. A 6 factor approximation algorithm was given by Khuller and Sussmann [15] for the case of uniform hard capacities (5 factor for soft capacitated case). For non-uniform hard capacities, Cygan *et al.* [13] gave the first constant approximation algorithm for the problem, which was further improved by An *et al.* in [3] to 9 factor.

Though the knapsack median problem (a.k.a. weighted W-median) is a well motivated problem and occurs naturally in practice, not much work has been done on the problem. Krishnaswamy *et al.* [17] showed that the integrality gap, for the uncapacitated case, holds even on adding the covering inequalities to strengthen the LP, and gave a 16 factor approximation that violates the budget constraint by a factor of $(1+\epsilon)$. Kumar [19] strengthened the natural LP by obtaining a bound on the maximum distance a client can travel and gave first constant factor approximation without violating the budget constraint. Charikar and Li [12] reduced the large constant obtained by Kumar to 34 which was further improved to 32 by Swamy [26]. Byrka *et al.* [7] extended the work of Swamy and applied sparsification as a pre-processing step to obtain a factor of 17.46. The result was further improed to $7.081(1+\epsilon)$ very recently by Krishnaswamy *et al.* [18] using iterative rounding technique, with a running time of $n^{O(1/\epsilon^2)}$.

For CkFLP, Aardal *et al.* [1] extended the FPTAS for knapsack problem to give an FPTAS for single client CkFLP. They also extend an $\alpha-$ approximation algorithm for (uncapacitated) k-median to give a $(2\alpha+1)-$ approximation for CkFLP with uniform opening costs using at most $2k$ for non-uniform and $2k-1$ for uniform capacities. Byrka *et al.* [6] gave an $O(1/\epsilon^2)$ factor approximation violating the capacities by a factor of $(2+\epsilon)$ using dependent rounding.

For CFLP, An, Singh and Svensson [4] gave the first LP-based constant factor approximation by strengthening the natural LP. Other LP-based algorithms known for the problem are due to Byrka *et al.* and Levi *et al.* ([6, 20]). The local search technique has been particularly useful to deal with capacities. The approach provides 3 factor for uniform capacities [2] and 5 factor for the non-uniform case [5].

1.3 Our techniques

We extend the work of Krishnaswamy *et al.* [17] to capacitated case. The major challenge is in writing the LP which opens sufficient number of facilities for us in bounded cost.

Filtering and clustering techniques [24, 11, 20, 25, 6, 17, 1] are used to partition the set of facilities and demands. Routing trees are used to bound the assignment costs. Main contribution of this work is a new LP and an iterative rounding algorithm to obtain a solution with at most two fractionally opened facilities.

High Level Ideas: We first use the filtering and clustering techniques to partition the set of facilities and demands. Each partition (*called cluster*) has sufficient opening $(\geq 1-1/\ell \geq 1/2)$ for a fixed parameter $\ell \geq 2$ in it. An integrally open solution is obtained where-in some clusters have at least 1 integrally opened facility and some do not have any facility opened in them. To assign the demand of the cluster that cannot be satisfied locally within the cluster, a (directed) rooted binary routing tree is constructed, on the cluster centers. If (s,t) is an edge in the routing tree then the cost of sending the unmet demand of the cluster centered at s to t is bounded. The edges of the tree have non-increasing costs as we go up the tree, with the root being at the top. Hence the cost of sending the unmet demand of the cluster centered at s to any node r up in the tree at a constant number of edges away from s is bounded.

In order to decide which facilities to open integrally, clusters are grouped into meta-clusters of size (the number of clusters in it) ℓ so as to have at least $\ell - 1$ opening in it. The routing tree is used to group the clusters into meta-clusters (MCs) in a top-down greedy manner, i.e., starting from the root, a meta-cluster grows by including the cluster (center) that connects to it by the cheapest edge. A MC grows until its size reaches ℓ. We then proceed to make a new MC from the tree with the remaining nodes in the same greedy manner. This imposes a natural directed (not necessarily binary) rooted tree structure on the meta-clusters with the property that the edge going out of a MC is cheaper than the edges inside the MC which are further cheaper than the edges coming into the MC. Out-degree of a MC is 1 whereas the in-degree is at most $q+1$ where q is the number of clusters in a MC.

Next, we write a new LP to open sufficient number of facilities within each cluster and each MC. We also give an iterative rounding algorithm to solve the LP, removing the integral variables and updating the constraints accordingly in each iteration until either all the variables are fractional or all are integral. In case all the variables are fractional, we use the property of extreme point solutions to claim that the number of non-integral variables is at most two. Thus we obtain a solution to the LP with at most two fractional openings. Both the fractionally opened facilities are opened integrally at a loss of additive f_{max} in the budget where f_{max} is the maximum facility opening cost [3].

Finally a min-cost flow problem is solved with capacities scaled up by a factor of $(2+\epsilon)$ to obtain an integral assignment. A feasible solution to the min-cost flow problem of bounded cost is obtained as follows: consider a scenario in which the demand accumulated within each cluster is less than u (we call such clusters as *sparse*). For the sake of easy exposition of the ideas, let each MC be of size exactly ℓ. The LP solution opens at least $\ell - 1$ facilities integrally in each MC, with at least one facility in each cluster except for one cluster. If the cluster with unmet demand is at the root of the induced subgraph of the MC, then its demand cannot be met within the MC. We make sure that such a demand is served in the parent MC. Total demand to be served by the facilities in a MC is at most ℓu plus at most $(\ell+1)u$ coming from the children of the MC. Thus $(\ell-1)$ facilities have to serve at most $(2\ell+1)u$ demand leading to a violation of $(2+O(1/\ell))$ in capacity. Demands have to travel $O(\ell)$ edges upwards (at most ℓ within its own MC and at most ℓ in the parent MC), and hence the cost of serving them is bounded.

The situation becomes a little tricky when there are clusters with more than u demand (we call such clusters as *dense*). One way to deal with dense clusters is to open $\lfloor demand/u \rfloor$ facilities integrally within such a cluster and assign the residual demand to one of them at a capacity violation of 2. But if this cluster also has to serve u units of unmet demand of one of its children (we will see later that a dense cluster has at most one child), the capacity violation could blow upto 3 in case $\lfloor demand/u \rfloor = 1$. We deal with this scenario carefully.

[3] Let F' be the set of facilities i with $f_i > \epsilon \cdot \mathcal{B}$. Enumerate all possible subsets of F' of $size <= 1/\epsilon$. There are at most $n^{O(1/\epsilon)}$ such sets. For each such set S, solve the LP with $y_i = 1 \; \forall \; i \in S$ and $y_i = 0 \; \forall \; i \in F' \setminus S$. The additive f_{max} (which comes from the fractionally opened facilities) is $<= \epsilon \cdot \mathcal{B}$. Choose the best solution and hence theorem 1 follows.

2 Capacitated Knapsack Median Problem

In this section, we consider the capacitated knapsack median problem. CKnM can be formulated as the following integer program (IP):

$$
\begin{aligned}
&\mathit{Minimize} \quad \mathit{CostKnM}(x,y) = \textstyle\sum_{j\in\mathcal{C}}\sum_{i\in\mathcal{F}} c(i,\ j)x_{ij} \\
&\mathit{subject\ to} \quad \textstyle\sum_{i\in\mathcal{F}} x_{ij} = 1 && \forall\ j\in\mathcal{C} && (1)\\
&\qquad \textstyle\sum_{j\in\mathcal{C}} x_{ij} \le u\ y_i && \forall\ i\in\mathcal{F} && (2)\\
&\qquad x_{ij} \le y_i && \forall\ i\in\mathcal{F},\ j\in\mathcal{C} && (3)\\
&\qquad \textstyle\sum_{i\in\mathcal{F}} f_i y_i \le \mathcal{B} && && (4)\\
&\qquad y_i,\ x_{ij} \in \{0,1\} && && (5)
\end{aligned}
$$

LP-Relaxation of the problem is obtained by allowing the variables $y_i, x_{ij} \in [0,1]$. Call it LP_1. To begin with, we guess the facility with maximum opening cost, f^*_{max}, in the optimal solution and remove all the facilities with facility cost $> f^*_{max}$ before applying the algorithm. For the easy exposition of ideas, we will give a weaker result, in section 2.4, in which we violate capacities by a factor of 3. Most of the ideas are captured in this section.

2.1 Simplifying the problem instance

We first simplify the problem instance by partitioning the sets of facilities and clients into clusters. This is achieved using the filtering technique of Lin and Vitter [24]. For an LP solution $\rho =< x, y >$ and a subset T of facilities, let $size(y,\ T) = \sum_{i\in T} y_i$ denote the total extent up to which facilities are opened in T under ρ.

Partitioning the set of facilities into clusters and sparsifying the client set : Let $\rho^* =< x^*, y^* >$ denote the optimal LP solution. Let $\hat{C}_j$ denote the average connection cost of a client j in ρ^* i.e., $\hat{C}_j = \sum_{i\in\mathcal{F}} x^*_{ij} c(i,\ j)$. Let $\ell \ge 2$ be a fixed parameter and *ball(j)* be the set of facilities within a distance of $\ell\hat{C}_j$ of j *i.e.*, $ball(j) = \{i \in \mathcal{F} \colon c(i,\ j) \le \ell\hat{C}_j\}$ (Figure 1(a)). Then, $size(y^*,\ ball(j)) \ge 1 - \frac{1}{\ell}$. Let $\mathcal{R}_j = \ell\hat{C}_j$ denote the *radius* of *ball(j)*. We identify a set $\mathcal{C}'$ of clients (Figure 1(b)) which will serve as the centers of the clusters using Algorithm 1. Note that $ball(j') \subseteq \mathcal{N}_{j'}$ and the sets $\mathcal{N}_{j'}$ partition $\mathcal{F}$. (Figure 2(b)).

Algorithm 1 Cluster Formation

1: $\mathcal{C}' \leftarrow \emptyset$, $S \leftarrow \mathcal{C}$, $ctr(j) = \emptyset\ \forall j \in S$.
2: **while** $S \ne \emptyset$ **do**
3: Pick $j' \in S$ with the smallest radius $\mathcal{R}_{j'}$ in S, breaking ties arbitrarily.
4: $S \leftarrow S \setminus \{j'\}$, $\mathcal{C}' \leftarrow \mathcal{C}' \cup \{j'\}$
5: **while** $\exists j \in S$: $c(j',\ j) \le 2\ell\hat{C}_j$ **do**
6: $S \leftarrow S \setminus \{j\}$, $ctr(j) = j'$
7: **end while**
8: **end while**
9: $\forall j' \in \mathcal{C}'$: let $\mathcal{N}_{j'} = \{i \in \mathcal{F} \mid \forall k' \in \mathcal{C}' : j' \ne k' \Rightarrow c(i,\ j') < c(i,\ k')\}$

Partitioning the demands: Let l_i denote the total demand of clients in $\mathcal{C}$ serviced by facility i i.e., $l_i = \sum_{j\in\mathcal{C}} x^*_{ij}$ and, $d_{j'} = \sum_{i\in\mathcal{N}_{j'}} l_i$ for $j' \in \mathcal{C}'$. Move the demand $d_{j'}$ to the center j' of the cluster (Figures 1-(b) and 2-(a)). For $j \in \mathcal{C}$, let $\mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'})$ denote the total extent upto which j is served by the facilities in $\mathcal{N}_{j'}$. Then, we can also write $d_{j'} = \sum_{j\in\mathcal{C}} \mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'})$. Thus, after this step, unit demand of any $j \in \mathcal{C}$, is distributed to centers of all the clusters whose facilities serve j. In particular, it takes care of the demand of the clients that were removed during sparsification. Each cluster center is then responsible for the portion of demand of $j \in \mathcal{C}$ served by the facilities in its cluster.

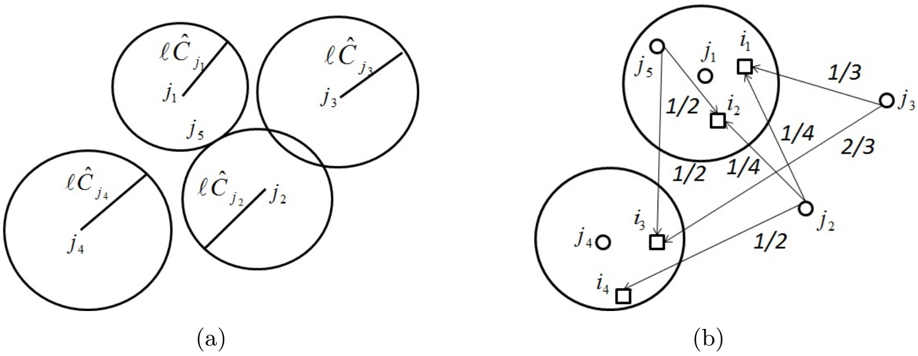

Figure 1 (a) The balls around the clients. (b) Reduced set of clients and assignment by LP solution.

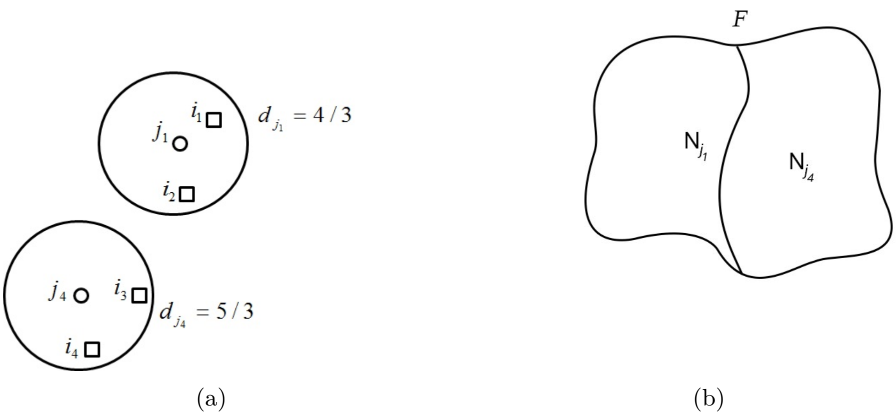

Figure 2 (a) Partitioning of demand. (b) Partition of $\mathcal{F}$.

The cost of moving the demand $d_{j'}$ to j' is bounded by $2(\ell+1)LP_{opt}$ as shown in Corollary 6. Also, any two cluster centers j' and k' satisfy the *separation property*: $c(j',\ k') > 2\ell\ max\{\hat{C_{j'}}, \hat{C_{k'}}\}$. In addition, they satisfy Lemmas (5), (7) and (8).

▶ **Lemma 5.** *Let $j' \in \mathcal{C}'$ and $i \in \mathcal{N}_{j'}$ then, (i) For $k' \in \mathcal{C}'$, $c(j',\ k') \leq 2c(i,\ k')$, (ii) For $j \in \mathcal{C} \setminus \mathcal{C}'$, $c(j',\ j) \leq 2c(i,\ j) + 2\ell\hat{C}_j$ and (iii) For $j \in \mathcal{C}$, $c(i,\ j') \leq c(i,\ j) + 2\ell\hat{C}_j$.*

Proof. $i)$ By triangle inequality, $c(j', k') \leq c(i, j') + c(i, k')$. Since $i \in \mathcal{N}_{j'} \Rightarrow c(i, j') \leq c(i, k')$ and hence $c(j', k') \leq 2c(i, k')$.

(ii) Since $j \notin \mathcal{C}'$, there exist a client $k' \in \mathcal{C}'$ such that $ctr(j) = k'$ and $c(j, k') \leq 2\ell\hat{C}_j$. Also, If $k' = j'$ then $c(i, j') = c(i, k')$ else $c(i, j') \leq c(i, k')$ because $i \in \mathcal{N}_{j'}$ and not $\mathcal{N}_{k'}$. Then, by triangle inequality, $c(i, k') \leq c(i, j) + c(j, k') \leq c(i, j) + 2\ell\hat{C}_j = c(i, j) + 2\mathcal{R}_j$. Therefore, $c(j', j) \leq c(i, j') + c(i, j) \leq 2c(i, j) + 2\mathcal{R}_j$.

(iii) Consider two cases: $j \in \mathcal{C}'$ and $j \notin \mathcal{C}'$. In the first case, $c(i, j') \leq c(i, j)$ because $i \in \mathcal{N}_{j'}$ and not $\mathcal{N}_j$ and hence $c(i, j') \leq c(i, j) + 2\ell\hat{C}_j$. In the latter case, by triangle inequality we have, $c(i, j') \leq c(i, j) + c(j', j)$. Since $j \notin \mathcal{C}' \Rightarrow c(j', j) \leq 2\ell\hat{C}_j$. Thus, $c(i, j') \leq c(i, j) + 2\ell\hat{C}_j$. ◀

▶ **Corollary 6.** $\sum_{j\in\mathcal{C}} \sum_{j'\in\mathcal{C}'} c(j',\ j)\mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'}) \leq 2(\ell+1)LP_{opt}$.

▶ **Lemma 7.** *Let $j \in \mathcal{C} \setminus \mathcal{C}'$ and $j' \in \mathcal{C}'$ such that $c(j',\ j) \leq \mathcal{R}_{j'}$, then $\mathcal{R}_{j'} \leq 2\mathcal{R}_j$.*

Proof. Suppose, if possible, $\mathcal{R}_{j'} > 2\mathcal{R}_j$. Let $ctr(j) = k'$. Then, $c(j,\ k') \leq 2\mathcal{R}_j$. And, $c(k',\ j') \leq c(k',\ j) + c(j,\ j') \leq 2\mathcal{R}_j + \mathcal{R}_{j'} < 2\mathcal{R}_{j'} = 2\ell\hat{C_{j'}}$, which is a contradiction to separation property. ◀

▶ **Lemma 8.** $\sum_{j' \in \mathcal{C}'} d_{j'} \sum_{i \in \mathcal{F}} c(i,\ j')x^*_{ij'} \leq 3 \sum_{j \in \mathcal{C}} \sum_{i \in \mathcal{F}} c(i,\ j)x^*_{ij} = 3LP_{opt}$.

Proof. $\sum_{j' \in \mathcal{C}'} d_{j'} \sum_{i \in \mathcal{F}} c(i,\ j')x^*_{ij'} = \sum_{j' \in \mathcal{C}'} \big(\sum_{j \in \mathcal{C}} \mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'})\big)\hat{C_{j'}}$
$= \sum_{j' \in \mathcal{C}'} \big(\sum_{j \in \mathcal{C}: c(j',j) \leq \mathcal{R}_{j'}} \mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'})\hat{C_{j'}} + \sum_{j \in \mathcal{C}: c(j',j) > \mathcal{R}_{j'}} \mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'})\hat{C_{j'}}\big)$
Second term in the sum on RHS $< \frac{1}{\ell} \sum_{j' \in \mathcal{C}'} \sum_{j \in \mathcal{C}: c(j',j) > \mathcal{R}_{j'}} \mathcal{A}_{\rho^*}(j, \mathcal{N}_{j'}) c(j',\ j)$
$\leq \frac{1}{\ell} \sum_{j \in \mathcal{C}} \sum_{j' \in \mathcal{C}': c(j',j) > \mathcal{R}_{j'}} \sum_{i \in \mathcal{N}_{j'}} x^*_{ij}(2c(i,\ j) + 2\ell\hat{C_j})$ as $c(j',\ j) \leq 2c(i,\ j) + 2\ell\hat{C_j}$ by Lemma 5
$\leq \sum_{j \in \mathcal{C}} \sum_{j' \in \mathcal{C}': c(j',j) > \mathcal{R}_{j'}} \sum_{i \in \mathcal{N}_{j'}} x^*_{ij}(c(i,\ j) + 2\hat{C_j})$. Thus the claim follows. ◀

Let $\mathcal{C}_S$ be the set of cluster centers $j' \in \mathcal{C}'$ for which $d_{j'} < u$ and $\mathcal{C}_D$ be the set of remaining centers in $\mathcal{C}'$. The clusters centered at $j' \in \mathcal{C}_S$ are called *sparse* and those centered at $j' \in \mathcal{C}_D$ *dense*. For $j' \in \mathcal{C}_D$, sufficient facilities are opened in $\mathcal{N}_{j'}$ so that its entire demand is served within the cluster itself and we say that j' is *self-sufficient*. Unfortunately, the same claim cannot be made for the sparse clusters i.e., we cannot guarantee to open even one facility in each sparse cluster (since $d_{j'} < u$, we need only one facility in each sparse cluster j'). Thus, in the next section, we define a routing tree that is used to route the unmet demand of a cluster to another cluster in bounded cost.

2.2 Constructing the Binary Routing Tree

First, we define a dependency graph $G = (V, E)$, similar to the one defined by Krishnaswamy et al [17], on cluster centers, i.e., $V = \mathcal{C}'$. For brevity of notation, we use j' to refer to the node corresponding to cluster center j' as well as to refer to the cluster center j' itself. For $j' \in \mathcal{C}_S$, let $\eta(j')$ be the nearest other cluster center in $\mathcal{C}'$ of j' i.e., $\eta(j') = k'(\neq j') \in \mathcal{C}' : k'' \in \mathcal{C}' \Rightarrow c(j',\ k') \leq c(j',\ k'')$ and for $j' \in \mathcal{C}_D,\ \eta(j') = j'$. The dependency graph consists of directed edges $c(j',\ \eta(j'))$. Each connected component of the graph is a tree except possibly for a 2-cycle at the root. We remove any edge arbitrarily from the two cycle. The resulting graph is then a forest. Note that, there is at most one dense cluster in a component and if present, it must be the root of the tree. The following lemma will be useful to bound the cost of sending the unserved demand of $j' \in \mathcal{C}_S$ to $\eta(j')$.

▶ **Lemma 9.** $\sum_{j' \in \mathcal{C}_S} d_{j'} (\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')x^*_{ij'} + c(j',\ \eta(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} x^*_{ij'})) \leq 6LP_{opt}$.

Proof. The second term of LHS $= \sum_{j' \in \mathcal{C}_S} d_{j'} \big(\sum_{i \notin \mathcal{N}_{j'}} c(j',\ \eta(j'))x^*_{ij'}\big)$
$\leq \sum_{j' \in \mathcal{C}_S} d_{j'} \big(\sum_{k' \in \mathcal{C}': k' \neq j'} \sum_{i \in \mathcal{N}_{k'}} c(j',\ k')x^*_{ij'}\big)$
$\leq \sum_{j' \in \mathcal{C}_S} d_{j'} \big(\sum_{k' \in \mathcal{C}': k' \neq j'} \sum_{i \in \mathcal{N}_{k'}} 2c(i,\ j')x^*_{ij'}\big)$. ◀

Unfortunately, the in-degree of a node in a tree may be unbounded and hence arbitrarily large amount of demand may accumulate at a cluster center, which may further lead to unbounded capacity violation at the facilities in its cluster.

Bounding the in-degree of a node in the dependency graph: We convert the dependency graph G into another graph G' where-in the in-degree of each node is bounded by 2 with in-degree of the root being 1. This is done as follows (Figure 3(a)-(b)): let $\mathcal{T}$ be a tree in G. $\mathcal{T}$ is converted into a binary tree using the standard procedure after sorting

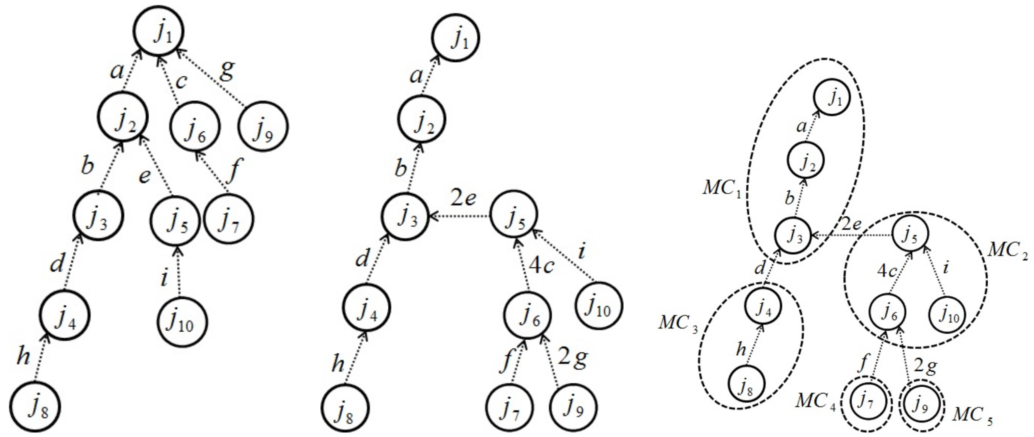

Figure 3 (a) A Tree T of unbounded in-degree. $a < b < d < h$, $a < c < g$, $b < e$. (b) A Binary Tree T' where each node has in-degree at most 2. (c) Formation of meta-clusters for $\ell = 3$.

the children of node j' from left to right in non-decreasing order of distance from j' i.e., for each child k' (except for the nearest child) of j', add an edge to its left sibling with weight $2c(k',\ \eta(k'))$ and remove the edge (k', j'). There is no change in the outgoing edge of the leftmost child of j'. Let $\psi(j')$ be the parent of node j' in G'. Its easy to see that $c(j',\ \psi(j')) \leq 2c(j',\ \eta(j'))$. Henceforth whenever we refer to distances, we mean the new edge weights. Hence, we have the following:

$$\sum_{j' \in \mathcal{C}_S} d_{j'} \Big(\sum_{i \in \mathcal{N}_{j'}} c(i,\ j') x^*_{ij'} + c(j',\ \psi(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} x^*_{ij'}) \Big) \leq 12 LP_{opt} \qquad (6)$$

2.3 Constructing the Meta-clusters

If we could ensure that for every $j' \in \mathcal{C}_S$ for which no facility is opened in $\mathcal{N}_{j'}$, a facility is opened in $\psi(j')$, we are done (with 3 factor loss in capacities). But we do not know how to do that. However, for every such cluster center j', we will identify a set of centers which will be able to take care of the demand of j' and each one of them is within a distance of $O(\ell)c(j',\ \psi(j'))$ from j'.

We exploit the following observation to make groups of ℓ clusters: each cluster has facilities opened in it to an extent of at least $(1 - 1/\ell)$. Hence, every collection of ℓ clusters, has at least $\ell - 1$ facilities opened in it. Thus, we make groups (called meta-clusters), each consisting of ℓ clusters, if possible. For every tree $\mathcal{T}$ in G', MCs are formed by processing the nodes of $\mathcal{T}$ in a top-down greedy manner starting from the root as described in Algorithm 2. (Also see Figure 3(c)). There may be some MCs of size less than ℓ, towards the leaves of the tree.

Let G_r denote a MC with r being the root cluster of it. With a slight abuse of notation, we will use G_r to denote the collection of centers of the clusters in it as well as the set of clusters themselves. Let $\mathcal{H}(G_r)$ denote the subgraph of $\mathcal{T}$ induced by the nodes in G_r. $\mathcal{H}(G_r)$ is clearly a tree. We say that G_r is responsible for serving the demand in its clusters.

With the guarantee of only $\ell - 1$ opening amongst ℓ clusters, there may be a cluster with no facility opened in it. If this cluster happens to be a sparse cluster at the root, its demand cannot be served within the MC. Thus we define a (routing) tree structure on MCs as follows: a tree consists of MCs as nodes and there is an edge from a MC G_r to another

Algorithm 2 Meta-cluster Formation

1: **Meta-cluster(Tree $\mathcal{T}$)**
2: $\mathcal{N} \leftarrow$ *set of nodes in* $\mathcal{T}$.
3: **while** there are non-grouped nodes in $\mathcal{N}$ **do**
4: Pick a topmost non-grouped node, say k of $\mathcal{N}$: form a new MC, G_k.
5: **while** G_k has fewer than ℓ nodes **do**
6: If $\mathcal{N} = \emptyset$ then break and stop.
7: Let $j = argmin_{u \in \mathcal{N}}\{c(u,\ v) : (u, v) \in \mathcal{T}, v \in G_k\}$, set $G_k = G_k \cup \{j\}$. $\mathcal{N} \leftarrow \mathcal{N} \setminus \{j\}$.
8: **end while**
9: **end while**

MC G_s if there is a directed edge from root r of G_r to some node $s' \in G_s$, G_s is then called the parent meta-cluster of G_r, G_r a child meta-cluster of G_s and the edge $(r,\ s')$ is called the *connecting edge* of the child MC G_r. If G_r is a root MC, add an edge to itself with cost $c(r,\ \ \psi(r))$. This edge is then called the *connecting edge* of G_r. Note that the cost of any edge in G_s is less than the cost of the connecting edge of G_r which is further less than the cost of any edge in G_r. Further, a dense cluster, if present, is always the root cluster of a root MC. We guarantee that the unmet demand of a MC is served in its parent MC.

2.4 3-factor capacity violation

In this section, we present the main contribution of our work. Inspired by the LP of Krishnaswamy *et al.* [17], we formulate a new LP and present an iterative rounding algorithm to obtain a solution with at most two fractionally opened facilities. Such a solution is called *pseudo-integral* solution. Modifying the LP of Krishnaswamy *et al.* [17] and obtaining a feasible solution of bounded cost for the capacitated scenario is non-trivial. The rounding algorithm is also non-trivial.

2.4.1 Formulating the new LP and obtaining a pseudo-integral solution

Sparse clusters have the nice property that they need to take care of small demand ($< u$ each) and dense clusters have the nice property that the total opening within each cluster is at least 1. These properties are exploited to define a new LP that opens sufficient number of facilities in each MC such that the opened facilities are well spread out amongst the clusters (we make sure that at most 1 (sparse) cluster has no facility opened in it) and demand of a dense cluster is satisfied within the cluster itself. We then present an iterative rounding algorithm that provides us with a solution having at most two fractionally opened facilities.

Let δ_r be the number of dense clusters and σ_r be the number of sparse clusters in a MC G_r. With at least $1 - 1/\ell$ opening in each sparse cluster, observing the fact that $\sigma_r \leq \ell$, we have at least $\sigma_r(1 - 1/\ell) \geq \sigma_r - 1$ total opening in σ_r sparse clusters of G_r. Also, at least $\lfloor d_{j_d}/u \rfloor$ opening is there in a dense cluster centered at j_d in G_r. Let $\alpha_r = \max\{0, \sigma_r - 1\}$. LP is defined so as to open at least $\lfloor d_{j_d}/u \rfloor + \alpha_r$ facilities in G_r. Let $\tau(j') = \{i \in \mathcal{N}_{j'} : c(i,\ j') \leq c(j',\ \psi(j'))\}$ if $j' \in \mathcal{C}_S$ (recall that $\psi(j')$ is the parent of j' in binary tree) and $\tau(j') = \mathcal{N}_{j'}$ if $j' \in \mathcal{C}_D$. Also, let $\mathcal{S}_r = G_r \cap \mathcal{C}_S$ and $s_r = \alpha_r$ for all MCs G_r, $\tilde{\mathcal{F}} = \mathcal{F}$, $\tilde{\mathcal{B}} = \mathcal{B}$, $r_{j'} = \lfloor d_{j'}/u \rfloor$ $\forall j' \in \mathcal{C}_D$ and $\hat{\tau}(j') = \tau(j')$ $\forall j' \in \mathcal{C}'$. These sets are updated as we go from one iteration to the next iteration in our rounding algorithm, thereby giving a new (reduced) LP in each iteration. Let w_i denote whether facility i is opened in the solution or not. We now write an LP, called LP_2 with the objective of minimising the following function:

$$CostKM(w) = \sum_{j' \in \mathcal{C}_S} d_{j'}[\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w_i + c(j',\ \psi(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} w_i)] + u \sum_{j' \in \mathcal{C}_D} \sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w_i$$

$$s.t. \quad \textstyle\sum_{i \in \hat{\tau}(j')} w_i \leq 1 \qquad \forall\ j' \in \mathcal{C}_S \tag{7}$$

$$\textstyle\sum_{i \in \hat{\tau}(j')} w_i = r_{j'} \qquad \forall\ j' \in \mathcal{C}_D \tag{8}$$

$$\textstyle\sum_{j' \in \mathcal{S}_r} \sum_{i \in \hat{\tau}(j')} w_i \geq s_r \quad \forall\ r : G_r \text{ is a MC} \tag{9}$$

$$\textstyle\sum_{i \in \tilde{\mathcal{F}}} f_i w_i \leq \tilde{\mathcal{B}} \tag{10}$$

$$0 \leq w_i \leq 1\ \forall\ i \in \tilde{\mathcal{F}} \tag{11}$$

Constraints (8) and (9) ensure that sufficient number of facilities are opened in a meta-cluster. Constraints (7) and (8) ensure that the opened facilities are well spread out amongst the clusters as no more than 1 and $\lfloor \frac{d_{j'}}{u} \rfloor$ facilities are opened in a sparse and dense cluster respectively. Constraint (8) also ensures that at least $\lfloor \frac{d_{j'}}{u} \rfloor$ facilities are opened in a dense cluster. This requirement is essential to make sure that the demand of a dense cluster is served within the cluster only. Hence, equality in constraint (8) is important.

▶ **Lemma 10.** *A feasible solution w' to LP_2 can be obtained such that $CostKM(w') \leq (2\ell + 13)LP_{opt}$.*

Proof. Refer to Appendix 5.1. ◀

For a vector $w \in \mathcal{R}^{|\mathcal{F}|}$ and $\mathcal{F}' \subseteq \mathcal{F}$, let $w^{\mathcal{F}'}$ denote the vector 'w restricted to $\mathcal{F}'$'. Also, let $s =< s_r >$, $S =< S_r >$ and $R =< r_{j'} >_{j' \in \mathcal{C}_D}$. Algorithm 3 presents an iterative rounding algorithm that solves LP_2 and returns a pseudo-integral solution $\tilde{w}$. A sparse cluster is removed from the scenario for the next iteration as and when a facility is integrally opened in it (lines 11, 12). In a dense cluster centered at j', the number of facilities to be opened by the LP ($r_{j'}$) is decremented by the number of integrally opened facilities in it (line 15) at every iteration and the cluster is removed when it becomes 0 (line 16). Similar treatment is done for $G_r \cap \mathcal{C}_S$ (line 12, 14)

▶ **Lemma 11.** *The solution $\tilde{w}$ given by Iterative Rounding Algorithm satisfies the following: i) $\tilde{w}$ is feasible, ii) $\tilde{w}$ has at most two fractional facilities and iii) $CostKM(\tilde{w}) \leq (2\ell + 13)LP_{opt}$.*

Proof. Refer to Appendix 5.2. ◀

2.4.2 Obtaining an integrally open solution

The two fractionally opened facilities obtained in Section 2.4.1, if any, are opened integrally at a loss of additive f_{max} in the budget. Let $\hat{w}$ denote the solution obtained. Next lemma shows that $\hat{w}$ has sufficient number of facilities opened in each MC to serve the demand the MC is responsible for, except possibly for u units. Lemma (12) presents the assignments done within a MC and discusses their impact on the capacity and the cost bounds.

▶ **Lemma 12.** *Consider a meta-cluster G_r. Suppose the capacities are scaled up by a factor of $\max\{3, 2 + \frac{4}{\ell - 1}\}$ for $\ell \geq 2$. Then, i) the dense cluster in G_r (if any) is* self-sufficient *i.e., its demand can be completely assigned within the cluster itself at a loss of at most factor 2 in cost. ii) There is at most one cluster with no facility opened in it and it is a sparse cluster. iii) Any (cluster) center responsible for the unserved demand of $j' \in \mathcal{C}'$ is an ancestor of j' in $\mathcal{H}(G_r)$. iv) At most u units of demand in G_r remain un-assigned and it must be in the root cluster of G_r. Such a MC cannot be a root MC. v) Let $\beta_r = \lfloor d_{j_d}/u \rfloor + \max\{0, \sigma_r - 1\}$, where j_d is the center of the dense root cluster (if any) in G_r. Then, at least β_r facilities*

Algorithm 3 Obtaining a pseudo-integral solution

pseudo-integral($\tilde{\mathcal{F}}$, $\tilde{\mathcal{B}}$, s, S, $\hat{\tau}()$, R)
$\tilde{w}_i^{\mathcal{F}} = 0 \; \forall i \in \mathcal{F}$
while $\tilde{\mathcal{F}} \neq \emptyset$ **do**
  Compute an extreme point solution $\tilde{w}^{\tilde{\mathcal{F}}}$ to LP_2.
  $\tilde{\mathcal{F}}_0 \leftarrow \{i \in \tilde{\mathcal{F}} : \tilde{w}_i^{\tilde{\mathcal{F}}} = 0\}$, $\tilde{\mathcal{F}}_1 \leftarrow \{i \in \tilde{\mathcal{F}} : \tilde{w}_i^{\tilde{\mathcal{F}}} = 1\}$.
  if $|\tilde{\mathcal{F}}_0| = 0$ and $|\tilde{\mathcal{F}}_1| = 0$ **then**
    Return $\tilde{w}^{\mathcal{F}}$. * exit when all variables are fractionally opened*\
  else
    For all MCs G_r\{
    while $\exists j' \in S_r$ such that constraint (7) is tight over $\tilde{\mathcal{F}}_1$ i.e., $\sum_{i \in \hat{\tau}(j') \cap \tilde{\mathcal{F}}_1} \tilde{w}_i^{\tilde{\mathcal{F}}} = 1$ **do**
      Remove the constraint corresponding to j' from (7). * a facility in $\tau(j')$ has been opened*\
      set $S_r = S_r \setminus \{j'\}$, $s_r = \max\{0, s_r - 1\}$. * delete the contribution of j' in constraint (9)*\
    end while
    If $s_r = 0$, remove the constraint corresponding to $\mathcal{S}_r$ from (9). * $\sigma_r - 1$ facilities have been opened in $G_r \cap \mathcal{C}_S$ *\
    If $\exists j' \in G_r \cap \mathcal{C}_D$, set $r_{j'} \leftarrow r_{j'} - |\hat{\tau}(j') \cap \tilde{\mathcal{F}}_1|$. * decrement $r_{j'}$ by the number of integrally opened facilities in $\hat{\tau}(j')$ *\
    If $r_{j'} = 0$, remove the constraint corresponding to j' from (8). * $\lfloor d_{j'}/u \rfloor$ facilities have been integrally opened in $\tau(j')$ *\ \}
  end if
  $\tilde{\mathcal{F}} \leftarrow \tilde{\mathcal{F}} \setminus (\tilde{\mathcal{F}}_0 \cup \tilde{\mathcal{F}}_1)$, $\tilde{\mathcal{B}} \leftarrow \tilde{\mathcal{B}} - \sum_{i \in \tilde{\mathcal{F}}_1} f_i \tilde{w}_i^{\tilde{\mathcal{F}}}$, $\hat{\tau}(j') \leftarrow \hat{\tau}(j') \setminus (\tilde{\mathcal{F}}_1 \cup \tilde{\mathcal{F}}_0) \; \forall j' \in \mathcal{C}'$.
end while
Return $\tilde{w}^{\mathcal{F}}$

are opened in G_r. (vi) Total distance traveled by demand $d_{j'}$ of $j'(\neq r) \in G_r$ to reach the centers of the clusters in which they are served is bounded by $d_{j'} c(j', \; \psi(j'))$.

Proof. Refer to Appendix 5.3. ◀

Lemma (13) deals with the remaining demand that we fail to assign within a MC. Such demand is assigned in the parent MC. Lemma (13) discusses the cost bound for such assignments and the impact of the demand coming onto G_r from the children MCs along with the demand within G_r on capacity.

▶ **Lemma 13.** *Consider a meta-cluster G_r. The demand of G_r and the demand coming onto G_r from the children meta-clusters can be assigned to the facilities opened in G_r such that: i) capacities are violated at most by a factor of $max\{3, 2 + \frac{4}{\ell - 1}\}$ for $\ell \geq 2$. ii) Total distance traveled by demand $d_{j'}$ of $j' \in \mathcal{C}'$ to reach the centers of the clusters in which they are served is bounded by $\ell d_{j'} c(j', \; \psi(j'))$.*

Proof. Refer to Appendix 5.4. ◀

Choosing $\ell \geq 2$ such that $2 + \frac{4}{(\ell-1)} = 3 \Rightarrow \ell = 5$. Lemma (14) bounds the cost of assigning the demands collected at the centers to the facilities opened in their respective clusters.

▶ **Lemma 14.** *The cost of assigning the demands collected at the centers to the facilities opened in their respective clusters is bounded by $O(1)LP_{opt}$.*

Proof. The proof follows from the observation that if $d_{j'}$ is served by a facility in $\tau(j''), j'' \in \mathcal{C}_S$ then $c(j'', \; i) \leq c(j'', \; \psi(j'')) \leq c(j', \; \psi(j'))$. This was the motivation to define $\tau(j')$ the way it was, while defining LP_2. For details, refer to Appendix 5.5. ◀

2.5 $(2+\epsilon)$ factor capacity violation

There is only one scenario in which we violate the capacities by a factor of 3 in the previous section. In all other scenarios capacities scaled up by a factor of $(2+\epsilon)$ are sufficient even to accommodate the demand of the children MCs. Consider this special scenario. Let j_d be the center of the dense cluster and j_s be its only child (sparse) cluster in the routing tree. Further let, $d_{j_d} = 1.99u$ and $d_{j_s} = .99u$. Then, we must have a total opening of more than 2 in the clusters of j_d and j_s taken together whereas LP_2 opens only 1. In such a scenario, if we treat j_s with j_d instead of considering it with the remaining sparse clusters of G_r, we can open 2 facilities in $\tau(j_d) \cup \tau(j_s)$ and they have to serve a total demand of at most $4u$ ($1.99u + .99u+$ at most u of the remaining sparse clusters) within the MC, thereby violating the capacities by a factor of at most 2. On the other hand, if $d_{j_d} = 1.01u$ and $d_{j_s} = .98u$, then we cannot guarantee to open 2 facilities in $\tau(j_d) \cup \tau(j_s)$. In this case, if we treated j_s with j_d and only 1 facility is opened in $\tau(j_d) \cup \tau(j_s)$, it will have to serve a total demand of (close to) $3u$ ($1.01u + .98u+$ at most u of the remaining sparse clusters) leading to violation of 3 in capacity. Note that first case corresponds to the scenario when the residual demand of j_d (viz. $.99u$ here) is large (close to u) and the second case corresponds to the scenario when the residual demand of j_d (viz. $.01u$ here) is small (close to 0). In the first case we treat j_s with j_d whereas in the second case, we treat it with the remaining sparse clusters. In Section 2.4, one can imagine that a MC G_r is partitioned into G^1_r and G^2_r where G^1_r contained only the dense cluster of G_r and G^2_r contained all the sparse clusters of G_r. We modify the partitions as follows: let $res(j_d) = d_{j_d}/u - \lfloor d_{j_d}/u \rfloor$: (i) if $res(j_d) < \epsilon$: set $G^1_r = G_r \cap \mathcal{C}_D$, $G^2_r = G_r \cap \mathcal{C}_S$, $\gamma_r = \lfloor d_{j_d}/u \rfloor$, $\sigma'_r = \sigma_r$. (This is same as above.) (ii) otherwise, $\epsilon \le res(j_d) < 1$: set $G^1_r = (G_r \cap \mathcal{C}_D) \cup \{j_s\}$, $G^2_r = (G_r \cap \mathcal{C}_S) \setminus \{j_s\}$, $\gamma_r = \lfloor d_{j_d}/u \rfloor + |\{j_s\}|$ [4], $\sigma'_r = \max\{0, \sigma_r - 1\}$.

We modify our LP accordingly so as to open at least γ_r facilities in G^1_r and $\alpha_r = \max\{0, \sigma'_r - 1\}$ facilities in G^2_r. Let $S^1_r = G^1_r$, $s^1_r = \gamma_r$ and $S^2_r = G^2_r$, $s^2_r = \alpha_r$, $\hat{\tau}(j') = \tau(j')\ \forall j'$. For $j' \in \mathcal{C}_D$, let $r_{j'} = \lfloor d_{j'}/u \rfloor$. Also, let $\tilde{\mathcal{F}} = \mathcal{F}$ and $\tilde{\mathcal{B}} = \mathcal{B}$. Let w_i denote whether facility i is opened in the solution or not. LP_2 is modified as follows:

$LP_3 : Min.\ \mathcal{C}ostKM(w)$

$$subject\ to \sum_{i \in \hat{\tau}(j')} w_i \le 1 \quad \forall\, j' \in \mathcal{C}_S \tag{12}$$

$$\sum_{j' \in S^1_r} \sum_{i \in \hat{\tau}(j')} w_i \ge s^1_r \quad \forall\, G^1_r : s^1_r \neq 0 \tag{13}$$

$$\sum_{j' \in S^2_r} \sum_{i \in \hat{\tau}(j')} w_i \ge s^2_r \quad \forall\, G^2_r : s^2_r \neq 0 \tag{14}$$

$$\sum_{i \in \tilde{\mathcal{F}}} f_i w_i \le \tilde{\mathcal{B}} \tag{15}$$

$$0 \le w_i \le 1 \qquad \forall i \in \tilde{\mathcal{F}} \tag{16}$$

▶ **Lemma 15.** *A feasible solution w' to LP_3 can be obtained such that $\mathcal{C}ostKM(w') \le (2\ell + 13)LP_{opt}$.*

Proof. Proof is similar to the proof of Lemma (10). ◀

Algorithm 3 can be modified to obtain Algorithm 4 as follows: whenever a constraint corresponding to (12) gets tight over integrally opened facilities, it is removed from S^1_r or S^2_r wherever it belongs, in the same manner as line 12 of Algorithm 3.

[4] In case a component of dependency graph consists of a singleton dense cluster, j_s may not exist. This case causes no problem even if $res(j_d)$ is large as it must be a leaf MC in this case.

Algorithm 4 Obtaining a pseudo-integral solution

pseudo-integral($\tilde{\mathcal{F}}$, $\tilde{\mathcal{B}}$, s^1, s^2, S^1, S^2, $\hat{\tau}()$, R')
$\tilde{w}^{\mathcal{F}}_i = 0 \ \forall i \in \mathcal{F}$
while $\tilde{\mathcal{F}} \neq \emptyset$ **do**
  Compute an extreme point solution $\tilde{w}^{\tilde{\mathcal{F}}}$ to LP_3.
  $\tilde{\mathcal{F}}_0 \leftarrow \{i \in \tilde{\mathcal{F}} : \tilde{w}^{\tilde{\mathcal{F}}}_i = 0\}$, $\tilde{\mathcal{F}}_1 \leftarrow \{i \in \tilde{\mathcal{F}} : \tilde{w}^{\tilde{\mathcal{F}}}_i = 1\}$.
  if $|\tilde{\mathcal{F}}_0| = 0$ and $|\tilde{\mathcal{F}}_1| = 0$ **then**
    Return $\tilde{w}^{\mathcal{F}}$.
  else
    For all MCs G_r{
    while $\exists j' \in G_r \cap \mathcal{C}_S$ such that constraint (12) is tight over $\tilde{\mathcal{F}}_1$ i.e., $\sum_{i \in \hat{\tau}(j') \cap \tilde{\mathcal{F}}_1} \tilde{w}^{\tilde{\mathcal{F}}}_i = 1$ **do**
      Remove the constraint corresponding to j' from (12). * a facility in $\tau(j')$ has been opened*\
      If $j' \in S^1_r$, set $S^1_r = S^1_r \setminus \{j'\}$, $s^1_r = \max\{0, s^1_r - 1\}$. * delete the contribution of j' in constraint (13) *\
      If $j' \in S^2_r$, set $S^2_r = S^2_r \setminus \{j'\}$, $s^2_r = \max\{0, s^2_r - 1\}$.* delete the contribution of j' in constraint (14) *\
      If $s^2_r = 0$, remove the constraint corresponding to the MC from (14).* α_r facilities have been opened in $G_r \cap \mathcal{C}_S$ *\
    end while
    If $\exists j' \in G_r \cap \mathcal{C}_D$, set $s^1_r = s^1_r - |\hat{\tau}(j') \cap \tilde{\mathcal{F}}_1|$. * decrement s^1_r by the number of integrally opened facilities in $\hat{\tau}(j')$ *\
    If $s^1_r = 0$, remove the constraint corresponding to the MC from (13). * γ_r facilities have been opened in G^1_r *\
  end if
  $\tilde{\mathcal{F}} \leftarrow \tilde{\mathcal{F}} \setminus (\tilde{\mathcal{F}}_0 \cup \tilde{\mathcal{F}}_1)$, $\tilde{\mathcal{B}} \leftarrow \tilde{\mathcal{B}} - \sum_{i \in \tilde{\mathcal{F}}_1} f_i \tilde{w}^{\tilde{\mathcal{F}}}_i$, $\hat{\tau}(j') \leftarrow \hat{\tau}(j') \setminus (\tilde{\mathcal{F}}_1 \cup \tilde{\mathcal{F}}_0) \ \forall j' \in \mathcal{C}'$.
end while
Return $\tilde{w}^{\mathcal{F}}$.

▶ **Lemma 16.** *The solution $\tilde{w}$ given by Iterative Rounding Algorithm satisfies the following: i) $\tilde{w}$ is feasible, ii) $\tilde{w}$ has at most two fractional facilities and iii) $\mathcal{C}ostKM(\tilde{w}) \leq (2\ell + 13)LP_{opt}$.*

Proof. Proof is similar to the proof of Lemma (11). ◀

The two fractionally opened facilities, if any, are opened integrally as in Section 2.4.2 at a loss of additive f_{max} in the budget. Let $\hat{w}$ denote the integrally open solution.

In the next lemma, we show that $\hat{w}$ has sufficient number of facilities opened in each MC to serve the demand the MC is responsible for, except possibly for u units. Let M be the set of all meta clusters and M_1 be the set of meta clusters, each consisting of exactly one dense and one sparse cluster. MCs in M_1 need special treatment and will be considered separately. Lemma (17) presents the assignments done within a MC and discusses their impact on the capacity and the cost bounds.

▶ **Lemma 17.** *Consider a meta-cluster G_r. Suppose the capacities are scaled up by a factor of $2 + \epsilon$ for $\ell \geq 1/\epsilon$. Then, (i) G^1_r is* self-sufficient *i.e., its demand can be completely assigned within the cluster itself. (ii) There are at most two clusters, one in G^1_r and one in G^2_r, with no facility opened in them and these clusters are sparse. (iii) Any (cluster) center responsible for the unserved demand of j' is an ancestor of j' in $\mathcal{H}(G_r)$. (iv) At most u units of demand in G_r remain un-assigned and it must be in the root cluster of G_r. Such a MC cannot be a root MC. (v) For $G_r \in M \setminus M_1$, let $\beta_r = \lfloor d_{j_d}/u \rfloor + \max\{0, \sigma_r - 1\}$, where j_d is the center of the dense root cluster in G_r. Then, at least β_r facilities are opened in G_r. (vi) For $G_r \in M_1$, let $\beta_r = \lfloor d_{j_d}/u \rfloor$ if $res(j_d) < \epsilon$ and $= \lfloor d_{j_d}/u \rfloor + 1$ otherwise. Then, at least β_r facilities are opened in G_r. (vii) Total distance traveled by demand $d_{j'}$ of $j' (\neq r) \in G_r$ to reach the centers of the clusters in which they are served is bounded by $2d_{j'}c(j',\ \psi(j'))$.*

Proof. Refer to Appendix 5.6. ◀

Lemma (18) deals with the remaining demand that we fail to assign within the MC. Such demand is assigned in the parent MC. Lemma (18) discusses the cost bound for such assignments and the impact of the demand coming onto G_r from the children MCs along with the demand within G_r on capacity.

▶ **Lemma 18.** *Consider a meta-cluster G_r. The demand of G_r and the demand coming onto G_r from the children meta-clusters can be assigned to the facilities opened in G_r such that: (i) capacities are violated at most by a factor of $(2+\frac{4}{\ell-1})$ for $\ell \geq 1/\epsilon$ and, (ii) Total distance traveled by demand $d_{j'}$ of $j' \in \mathcal{C}'$ to reach the centers of the clusters in which they are served is bounded by $\ell d_{j'} c(j', \ \psi(j'))$.*

Proof. Proof is similar to the proof of Lemma (13). ◀

▶ **Lemma 19.** *The cost of assigning the demands collected at the centers to the facilities opened in their respective clusters is bounded by $(2+\epsilon)(2\ell+1)LP_{opt}$.*

Proof. Proof is similar to the proof of Lemma (14). ◀

3 Capacitated k Facility Location Problem

Standard LP-Relaxation of the CkFLP can be found in Aardal *et al.* [1]. When $f_i = 0$, the problem reduces to the k-median problem and when $k = |\mathcal{F}|$ it reduces to the facility location problem. Our techniques for CKnM provide similar results for CkFLP in a straight forward manner i.e., $O(1/\epsilon^2)$ factor approximation, violating the capacities by a factor of $(2+\epsilon)$ and cardinality by plus 1. The violation of cardinality can be avoided by opening the facility with larger opening integrally while converting a pseudo integral solution into an integrally open solution. Thus, we obtain Theorem 2.

Proof of Theorem 3: Let $\rho^* =< x^*, y^* >$ denote the optimal LP solution. For sparse clusters, we open the cheapest facility i^* in *ball(j)*, close all facilities in the cluster and shift their demands to i^*. Let $\hat{\rho} =< \hat{x}, \hat{y} >$ be the solution so obtained. It is easy to see that we loose at most a factor of 2 in cardinality, and $\mathcal{C}ost\text{k}FLP(\hat{x}, \hat{y})$ is within $O(1)LP_{opt}$.

To handle dense clusters, we introduce the notion of cluster instances. For each cluster center $j' \in \mathcal{C}_D$, let $b^f_{j'} = \sum_{i \in \mathcal{N}_{j'}} f_i y^*_i$ and $b^c_{j'} = \sum_{i \in \mathcal{N}_{j'}} \sum_{j \in \mathcal{C}} x^*_{ij}[c(i, \ j) + 4\hat{C}_j]$. We define a cluster instance $\mathcal{S}_{j'}(j', \ \mathcal{N}_{j'}, \ d_{j'}, \ b^c_{j'}, b^f_{j'})$ as follows: Minimize $Cost_{CI}(z) = \sum_{i \in \mathcal{N}_{j'}} (f_i + uc(i, \ j'))z_i$ s.t. $u \sum_{i \in \mathcal{N}_{j'}} z_i \geq d_{j'}$ and $z_i \in [0, 1]$. It can be shown that $z_i = \sum_{j \in \mathcal{C}} x^*_{ij}/u = l_i/u \leq y^*_i \ \forall i \in \mathcal{N}_{j'}$ is a feasible solution with cost at most $b^f_{j'} + b^c_{j'}$. An almost integral solution z' is obtained by arranging the fractionally opened facilities in z in non-decreasing order of $f_i + c(i, \ j')u$ and greedily transferring the total opening $size(z, \ \mathcal{N}_{j'})$ to them. Let $l'_i = z'_i u$. For a fixed $\epsilon > 0$, an integrally open solution $\hat{z}$ and assignment $\hat{l}$ (possibly fractional) is obtained as follows: let i_1 be the fractionally opened facility, if any. If $z'_{i_1} < \epsilon$, close i_1 and shift its demand to another integrally opened facility at a loss of factor $(1+\epsilon)$ in its capacity. Else $(z'_{i_1} \geq \epsilon)$, open i_1, at a loss of factor 2 in cardinality and $1/\epsilon$ in facility cost. The solution $\hat{z}$ satisfies the following: $\hat{l}_i \leq (1+\epsilon)\hat{z}_i u \ \forall i \in \mathcal{N}_{j'}$, $\sum_{i \in \mathcal{N}_{j'}} \hat{z}_i \leq 2 \sum_{i \in \mathcal{N}_{j'}} z'_i \ \forall j' \in \ \mathcal{C}_D$ and $Cost_{CI}(\hat{z}) \leq \max\{1/\epsilon, 1+\epsilon\} Cost_{CI}(\hat{z})$.

4 Conclusion

In this work, we presented the first constant factor approximation algorithm for uniform hard capacitated knapsack median problem violating the budget by a factor of $(1+\epsilon)$ and capacity by $(2+\epsilon)$. Two variety of results were presented for capacitated k-facility location problem with a trade-off between capacity and cardinality violation: an $O(1/\epsilon^2)$ factor approximation violating capacities by $(2+\epsilon)$ and a $O(1/\epsilon)$ factor approximation, violating the capacity by a factor of at most $(1+\epsilon)$ using at most $2k$ facilities. As a by-product, we also gave a constant factor approximation for uniform capacitated facility location at a loss of $(1+\epsilon)$ in capacity from the natural LP. The result shows that the natural LP is not too bad.

It would be interesting to see if the capacity violation can be reduced to $(1+\epsilon)$ using the techniques of Byrka *et al.* [8]. Avoiding violation of budget will require strengthening the LP in a non-trivial way. Another direction for future work would be to extend our results to non-uniform capacities. Conflicting requirement of facility costs and capacities makes the problem challenging.

References

1. Karen Aardal, Pieter L. van den Berg, Dion Gijswijt, and Shanfei Li. Approximation algorithms for hard capacitated k-facility location problems. *EJOR*, 242(2):358–368, 2015. doi:10.1016/j.ejor.2014.10.011.
2. Ankit Aggarwal, Anand Louis, Manisha Bansal, Naveen Garg, Neelima Gupta, Shubham Gupta, and Surabhi Jain. A 3-approximation algorithm for the facility location problem with uniform capacities. *Journal of Mathematical Programming*, 141(1-2):527–547, 2013.
3. Hyung-Chan An, Aditya Bhaskara, Chandra Chekuri, Shalmoli Gupta, Vivek Madan, and Ola Svensson. Centrality of trees for capacitated k-center. *Math. Program.*, 154(1-2):29–53, 2015. URL: https://doi.org/10.1007/s10107-014-0857-y, doi:10.1007/s10107-014-0857-y.
4. Hyung-Chan An, Mohit Singh, and Ola Svensson. Lp-based algorithms for capacitated facility location. In *FOCS, 2014*, pages 256–265.
5. Manisha Bansal, Naveen Garg, and Neelima Gupta. A 5-approximation for capacitated facility location. In *Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia, September 10-12, 2012. Proceedings*, pages 133–144, 2012. doi:10.1007/978-3-642-33090-2_13.
6. Jaroslaw Byrka, Krzysztof Fleszar, Bartosz Rybicki, and Joachim Spoerhase. Bi-factor approximation algorithms for hard capacitated k-median problems. In *SODA 2015*, pages 722–736.
7. Jaroslaw Byrka, Thomas Pensyl, Bartosz Rybicki, Joachim Spoerhase, Aravind Srinivasan, and Khoa Trinh. An improved approximation algorithm for knapsack median using sparsification. In *ESA 2015*, pages 275–287.
8. Jaroslaw Byrka, Bartosz Rybicki, and Sumedha Uniyal. An approximation algorithm for uniform capacitated k-median problem with $(1+\epsilon)$ capacity violation. In *Integer Programming and Combinatorial Optimization - 18th International Conference, IPCO 2016, Liège, Belgium, June 1-3, 2016, Proceedings*, pages 262–274, 2016. doi:10.1007/978-3-319-33461-5_22.
9. Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS), New York, NY, USA*, pages 378–388, 1999.
10. Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for facility location problems. *SIAM Journal on Computing*, 34(4):803–824, 2005.
11. Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem (extended abstract). In *STOC, 1999*, pages 1–10. doi:10.1145/301250.301257.

12 Moses Charikar and Shi Li. A dependent lp-rounding approach for the k-median problem. In *ICALP 2012*, pages 194–205.

13 Marek Cygan, MohammadTaghi Hajiaghayi, and Samir Khuller. LP rounding for k-centers with non-uniform hard capacities. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 273–282, 2012. URL: `https://doi.org/10.1109/FOCS.2012.63`, `doi:10.1109/FOCS.2012.63`.

14 H. Gökalp Demirci and Shi Li. Constant approximation for capacitated k-median with $(1 + \epsilon)$ capacity violation. In *43rd ICALP 2016*, pages 73:1–73:14.

15 Samir Khuller and Yoram J. Sussmann. The capacitated K-center problem. *SIAM J. Discrete Math.*, 13(3):403–418, 2000. URL: `https://doi.org/10.1137/S0895480197329776`, `doi:10.1137/S0895480197329776`.

16 Madhukar R. Korupolu, C. Greg Plaxton, and Rajmohan Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37(1):146–188, 2000.

17 Ravishankar Krishnaswamy, Amit Kumar, Viswanath Nagarajan, Yogish Sabharwal, and Barna Saha. The matroid median problem. In *SODA, 2011*, pages 1117–1130.

18 Ravishankar Krishnaswamy, Shi Li, and Sai Sandeep. Constant approximation for k-median and k-means with outliers via iterative rounding. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 646–659, 2018. URL: `http://doi.acm.org/10.1145/3188745.3188882`, `doi:10.1145/3188745.3188882`.

19 Amit Kumar. Constant factor approximation algorithm for the knapsack median problem. In *SODA, 2012*, pages 824–832.

20 Retsef Levi, David B. Shmoys, and Chaitanya Swamy. Lp-based approximation algorithms for capacitated facility location. *Journal of Mathematical Programming*, 131(1-2):365–379, 2012.

21 Shanfei Li. An Improved Approximation Algorithm for the Hard Uniform Capacitated k-median Problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014), Germany*, pages 325–338.

22 Shi Li. On uniform capacitated k-median beyond the natural LP relaxation. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 696–707, 2015.

23 Shi Li. Approximating capacitated k-median with $(1 + \epsilon)k$ open facilities. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 786–796, 2016.

24 Jyh-Han Lin and Jeffrey Scott Vitter. epsilon-approximations with minimum packing constraint violation (extended abstract). In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 771–782, 1992. `doi:10.1145/129712.129787`.

25 David B. Shmoys, Éva Tardos, and Karen Aardal. Approximation algorithms for facility location problems (extended abstract). In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA,*, pages 265–274, 1997.

26 Chaitanya Swamy. Improved approximation algorithms for matroid and knapsack median problems and applications. In *APPROX/RANDOM, 2014*, pages 403–418, 2014.

5 Appendix

5.1 Proof of Lemma 10

Define a feasible solution to the LP_2 as follows: let $j' \in \mathcal{C}_D,\ i \in \tau(j')$, set $w'_i = \frac{l_i}{d_{j'}}\lfloor d_{j'}/u \rfloor = \frac{l_i}{u}\frac{\lfloor d_{j'}/u \rfloor}{d_{j'}/u} \leq \frac{l_i}{u} \leq y^*_i$. For $j' \in \mathcal{C}_S$, we set $w'_i = min\{x^*_{ij'},\ y^*_i\} = x^*_{ij'} \leq y^*_i$ for $i \in \tau(j')$ and $w'_i = 0$ for $i \in \mathcal{N}_{j'} \setminus \tau(j')$. We will next show that the solution is feasible.

For $j' \in \mathcal{C}_S$, $\displaystyle\sum_{i \in \tau(j')} w'_i \leq \sum_{i \in \mathcal{N}_{j'}} w'_i = \sum_{i \in \mathcal{N}_{j'}} x^*_{ij'} \leq 1$.

Next, let $j' \in \mathcal{C}_D$, then $\displaystyle\sum_{i \in \tau(j')} w'_i = \sum_{i \in \mathcal{N}_{j'}} \frac{l_i}{u}\frac{\lfloor d_{j'}/u \rfloor}{d_{j'}/u} = \lfloor d_{j'}/u \rfloor$ as $\displaystyle\sum_{i \in \mathcal{N}_{j'}} l_i = d_{j'}$. Note that $\displaystyle\sum_{i \in \tau(j')} w'_i \geq 1$ as $d_{j'} \geq u$.

For a meta-cluster G_r, we have $\displaystyle\sum_{j' \in G_r}\sum_{i \in \tau(j')} w'_i = \sum_{j' \in G_r \cap \mathcal{C}_S}\sum_{i \in \tau(j')} x^*_{ij'} \geq \sum_{j' \in G_r \cap \mathcal{C}_S}(1 - 1/l) = \max\{0, \sigma_r - 1\} = \alpha_r$.

Since for each $i \in \mathcal{F}$ we have $w'_i \leq y^*_i \Rightarrow \displaystyle\sum_{i \in \mathcal{F}} f_i w'_i \leq \sum_{i \in \mathcal{F}} f_i y^*_i \leq \mathcal{B}$.

Next, consider the objective function. For $j' \in \mathcal{C}_D$, we have $\displaystyle\sum_{i \in \tau(j')} u\ c(i,\ j')w'_i = u\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')(\frac{\sum_{j \in \mathcal{C}} x^*_{ij}}{u}) = \sum_{i \in \mathcal{N}_{j'}}\sum_{j \in \mathcal{C}} c(i,\ j')x^*_{ij} \leq \sum_{i \in \mathcal{N}_{j'}}\sum_{j \in \mathcal{C}}\left(c(i,\ j) + 2\ell\hat{C}_j\right)x^*_{ij}$. Summing over all $j' \in \mathcal{C}_D$ we get, $\displaystyle\sum_{j' \in \mathcal{C}_D}\sum_{i \in \mathcal{N}_{j'}}\sum_{j \in \mathcal{C}} x^*_{ij}[c(i,\ j) + 2\ell\hat{C}_j] \leq (2\ell + 1)LP_{opt}$.

Now consider the part of objective function for $\mathcal{C}_S$. $\sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w'_i + c(j',\ \psi(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} w'_i)) = \sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \tau(j')} c(i,\ j')w'_i + \sum_{i \in \mathcal{N}_{j'} \setminus \tau(j')} c(i,\ j')w'_i + c(j',\ \psi(j'))(1 - \sum_{i \in \tau(j')} w'_i - \sum_{i \in \mathcal{N}_{j'} \setminus \tau(j')} w'_i)) = \sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \tau(j')} c(i,\ j')x^*_{ij'} + c(j',\ \psi(j'))(1 - \sum_{i \in \tau(j')} x^*_{ij'}))$

$< \sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \tau(j')} c(i,\ j')x^*_{ij'} + c(j',\ \psi(j'))(1 - \sum_{i \in \tau(j')} x^*_{ij'})) + \sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \mathcal{N}_{j'} \setminus \tau(j')} (c(i,\ j') - c(j',\ \psi(j')))x^*_{ij'})$ as $c(i,\ j') > c(j',\ \psi(j'))\ \forall i \in \mathcal{N}_{j'} \setminus \tau(j')$

$= \sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')x^*_{ij'} + c(j',\ \psi(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} x^*_{ij'}))$. Thus, by equation (6), we get $\sum_{j' \in \mathcal{C}_S} d_{j'}(\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w'_i + c(j',\ \psi(j'))(1 - \sum_{i \in \mathcal{N}_{j'}} w'_i)) \leq 12LP_{opt}$.

Thus, the solution w' is feasible and $CostKM(w')$,

$$\sum_{j' \in \mathcal{C}_S} d_{j'}\left[\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w'_i + c(j',\ \psi(j'))\left(1 - \sum_{i \in \mathcal{N}_{j'}} w'_i\right)\right] + u\sum_{j' \in \mathcal{C}_D}\sum_{i \in \mathcal{N}_{j'}} c(i,\ j')w'_i \leq (2\ell + 13)LP_{opt}.$$

5.2 Proof of Lemma 11

$i)$ We will prove the claim by induction. Let $LP^{(t)}$ denote the LP at the beginning of the t^{th} iteration and $\tilde{w}^{(t)}$ denote the solution at the end of the t^{th} iteration. We will show that if $\tilde{w}^{(t)}$ is a feasible solution to LP_2, then $\tilde{w}^{(t+1)}$ is also a feasible solution to LP_2. Since $\tilde{w}^{(1)}$ is feasible (extreme point solution), the feasibility of the solution follows. Let $\tilde{\mathcal{F}}^{(t)}, \tilde{\mathcal{B}}^{(t)}, s^{(t)}, S^{(t)}, \hat{\tau}()^{(t)}, R^{(t)}$ denote the values at the beginning of the t^{th} iteration. Then, $\tilde{w}_i^{(t+1)} = \tilde{w}_i^{(t)}\ \forall i \in \mathcal{F} \setminus \tilde{\mathcal{F}}^{(t+1)}$.

Consider a constraint that was not present in $LP^{(t+1)}$. In any iteration, we remove a constraint only when none of the facilities in its corresponding clusters is fractionally opened. That is all the facilities in $\tau(j')$ appearing on the left hand side of a constraint are integral. Thus $\tilde{w}_i^{(t+1)} = \tilde{w}_i^{(t)}$ for all such facilities. Hence if they are satisfied by $\tilde{w}^{(t)}$ then they are satisfied by $\tilde{w}^{(t+1)}$. So, we consider only those constraints that were present in $LP^{(t+1)}$. For $j' \in \mathcal{C}_S$, since $\hat{\tau}(j')^{(t+1)} = \tau(j') \setminus \tilde{\mathcal{F}}_0^{(t)}\ \forall t$, therefore, $\sum_{i \in \hat{\tau}(j')^{(t+1)}} \tilde{w}_i^{(t+1)} = \sum_{i \in \tau(j')} \tilde{w}_i^{(t+1)}\ \forall t$. Thus, we will omit (t) and use $\tau()$ instead of $\hat{\tau}()$ for brevity of notation.

Consider constraints (7) that were not removed in t^{th} iteration. Since $\tau(j') \subseteq \tilde{\mathcal{F}}^{(t+1)}$ for $j' \in \mathcal{C}_S$, the feasibility of the constraint follows as $\tilde{w}^{(t+1)}$ is an extreme point solution of the reduced LP over the set $\tilde{\mathcal{F}}^{(t+1)}$.

Next, consider constraints (8). Let $\mathcal{F}_1^{(t)}$ denote the set of facilities that are opened integrally in $\tilde{w}^{(t)}$ i.e., $\tilde{w}_i^{(t)} = 1\ \forall i \in \mathcal{F}_1^{(t)}$ then the corresponding constraint in $LP^{(t+1)}$ is $\sum_{i \in \tau(j') \setminus \mathcal{F}_1^{(t)}} w_i = \lfloor \frac{d_{j'}}{u} \rfloor - |\mathcal{F}_1^{(t)}|$. Since $\tilde{w}^{(t+1)}$ is an extreme point solution of $LP^{(t+1)}$, it satisfies this constraint i.e., $\sum_{i \in \hat{\tau}(j') \setminus \mathcal{F}_1^{(t)}} \tilde{w}_i^{(t+1)} = \lfloor \frac{d_{j'}}{u} \rfloor - |\mathcal{F}_1^{(t)}|$. Since $w_i^{(t+1)} = w_i^{(t)} = 1\ \forall i \in \mathcal{F}_1^{(t)}$, adding $\mathcal{F}_1^{(t)}$ on both the sides, we get the desired feasibility.

Consider constraints (9). Since $\tilde{w}^{(t)}$ is feasible for LP_2, we have, $\sum_{j' \in G_r \cap \mathcal{C}_S} \sum_{i \in \tau(j')} \tilde{w}_i^{(t)} \geq \alpha_r$ and since $\tilde{w}^{(t+1)}$ is feasible for $LP^{(t+1)}$, we have $\sum_{j' \in S_r^{(t+1)}} \sum_{i \in \tau(j')} \tilde{w}_i^{(t+1)} \geq s_r^{(t+1)}$. Then, $\sum_{j' \in G_r \cap \mathcal{C}_S} \sum_{i \in \tau(j')} \tilde{w}_i^{(t+1)} = \sum_{j' \in (G_r \cap \mathcal{C}_S) \setminus S_r^{(t+1)}} \sum_{i \in \tau(j')} \tilde{w}_i^{(t+1)} + \sum_{j' \in S_r^{(t+1)}} \sum_{i \in \tau(j')} \tilde{w}_i^{(t+1)}$ $\geq \sum_{j' \in (G_r \cap \mathcal{C}_S) \setminus S_r^{(t+1)}} \sum_{i \in \tau(j')} \tilde{w}_i^{(t)} + s_r^{(t+1)} = \sum_{j' \in (G_r \cap \mathcal{C}_S) \setminus S_r^{(t+1)}} 1 + s_r^{(t+1)}$ (as these clusters must have been removed as they got tight) $= |(G_r \cap \mathcal{C}_S) \setminus S_r^{(t+1)}| + s_r^{(t+1)} = \alpha_r$

Next, consider constraint (10). Since $\tilde{w}^{(t)}$ is feasible for LP_2, we have $\sum_{i \in \mathcal{F}} f_i \tilde{w}_i^{(t)} \leq \mathcal{B}$ and since $\tilde{w}^{(t+1)}$ is feasible for $LP^{(t+1)}$, we have $\sum_{i \in \tilde{\mathcal{F}}^{(t+1)}} f_i \tilde{w}_i^{(t+1)} \leq \tilde{\mathcal{B}}^{(t+1)}$. Also, we have $w_i^{(t+1)} = w_i^{(t)}\ \forall i \in \mathcal{F} \setminus \tilde{\mathcal{F}}^{(t+1)}$. Consider $\sum_{i \in \mathcal{F}} f_i \tilde{w}_i^{(t+1)} = \sum_{i \in \mathcal{F} \setminus \tilde{\mathcal{F}}^{(t+1)}} f_i \tilde{w}_i^{(t+1)} + \sum_{i \in \tilde{\mathcal{F}}^{(t+1)}} f_i \tilde{w}_i^{(t+1)} \leq \sum_{i \in \mathcal{F} \setminus \tilde{\mathcal{F}}^{(t+1)}} f_i \tilde{w}_i^{(t)} + \tilde{\mathcal{B}}^{(t+1)}$. And since $\tilde{\mathcal{B}}^{(t+1)} = \mathcal{B} - \sum_{i \in \mathcal{F} \setminus \tilde{\mathcal{F}}^{(t+1)}} f_i \tilde{w}_i^{(t)}$, we have $\sum_{i \in \mathcal{F}} f_i \tilde{w}_i^{(t+1)} \leq \mathcal{B}$. Thus, the solution $\tilde{w}^{(t+1)}$ is feasible.

ii) Consider the last iteration of the algorithm. The iteration ends either at step $(3-4)$ or at step $(9-10)$. In the former case, the solution clearly has no fractionally opened facility. Suppose we are in the latter case. Let the linearly independent tight constraints corresponding to (7), (8) and (9) be denoted as $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ respectively. Let A and B be set of variables corresponding to some constraint in $\mathcal{X}$ and $\mathcal{Z}$ respectively such that $A \cap B \neq \emptyset$. Then, $A \subseteq B$. Imagine deleting A from B and subtracting 1 from s_r. Repeat the process with another such constraint in $\mathcal{X}$ until there is no more constraint in $\mathcal{X}$ whose variable set has a non-empty intersection with B. At this point, $s_r \geq 1$ and the number of variables in B is at least 2. Number of variables in any set corresponding to a tight constraint in $\mathcal{X}$ (or $\mathcal{Y}$) is also at least 2. Thus, the total number of variables is at least $2|\mathcal{X}| + 2|\mathcal{Y}| + 2|\mathcal{Z}|$ and the number of tight constraints is at most $|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| + 1$. Thus, we get $|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| \leq 1$ and hence there at most two (fractional) variables.

iii) Note that no facility is opened in $\mathcal{N}_{j'} \setminus \tau(j') : j' \in \mathcal{C}_S$ for if $i \in \mathcal{N}_{j'} \setminus \tau(j') : j' \in \mathcal{C}_S$ is opened, then it can be shut down and the demand $d_{j'} \tilde{w}_i$, can be shipped to $\psi(j')$, decreasing the cost as $c(j',\ \psi(j')) < c(i,\ j')$. Then, the claim follows as we compute extreme point solution in step (7) in the first iteration and the cost never increases in subsequent calls.

5.3 Proof of Lemma 12

(i) Let $j_d \in \mathcal{C}_D \cap G_r$. Total demand d_{j_d} of j_d can be distributed to the opened facilities ($\geq \lfloor d_{j_d}/u \rfloor$) at a loss of factor 2 in capacity and cost both, as $d_{j_d}/u - \lfloor d_{j_d}/u \rfloor < 1 \leq \lfloor d_{j_d}/u \rfloor$.

For $\sigma_r = 0$, (ii) - (v) hold vacuously. So, let $\sigma_r \geq 1$ (ii) LP_2 opens $\alpha_r = \max\{0, \sigma_r - 1\}$ facilities in $G_r \cap \mathcal{C}_S$. Constraint (7) ensures that at most one facility is opened in each sparse cluster. Thus, there is at most one cluster in $G_r \cap \mathcal{C}_S$ with no facility opened in it. (iii) & (iv) Let $j' \in G_r \cap \mathcal{C}_S$ such that no facility is opened in $\tau(j')$. If j' is not the root of G_r or G_r is a root MC, then LP_2 must have opened a facility in $\tau(\psi(j'))$. Demand of j' is assigned to this facility at a loss of maximum 2 factor in capacity if $\psi(j') \in \mathcal{C}_S$ and 3 if $\psi(j') \in \mathcal{C}_D$: $d_{\psi(j')} = 1.99u$ and $d_{j'} = .99u$. Otherwise (if j' is the root of G_r and G_r is not a root MC), at most u units of demand of G_r remain unassigned within G_r. (v) holds as$\lfloor d_{j_d}/u \rfloor$ facilities are opened in the cluster centered at j_d and $\alpha_r = \max\{0, \sigma_r - 1\}$ facilities are opened in $G_r \cap \mathcal{C}_S$ by constraints (8) and (9) respectively. (vi) Since the demand $d_{j'}$ of $j' \in G_r$ is served either within its own cluster or in the cluster centered at $\psi(j')$, total distance traveled by demand $d_{j'}$ of j' to reach the centers of the clusters in which they are served is bounded by $d_{j'} c(j',\ \psi(j'))$.

5.4 Proof of Lemma 13

After assigning the demands of the clusters within G_r as explained in Lemma (12), demand coming from all the children meta-clusters are distributed proportionately to facilities within G_r utilizing the remaining capacities. Next, we will show that this can be done within the claimed capacity bound.

(i) Let G_r be a non leaf meta-cluster with a dense cluster $j' \in \mathcal{C}_D$ at the root, if any. Also, let t_r be the total number of clusters in G_r, i.e., $t_r = \delta_r + \sigma_r$. The total demand to be served in G_r is at most $u(\lfloor d_{j'}/u \rfloor + 1 + \sigma_r) + u(t_r + 1) \leq (\beta_r + 2)u + (t_r + 1)u$ whereas the total available capacity is at least $\beta_r u$ by Lemma (12). Thus, the capacity violation is bounded by $\frac{(\beta_r+2)u+(t_r+1)u}{\beta_r u} \leq \frac{(\beta_r+2)u+(\beta_r+2)u}{\beta_r u} = 2 + 4/\beta_r \leq 2 + 4/(\ell - 1)$ (as $\lfloor d_{j'}/u \rfloor \geq \delta_r$ we have $\beta_r \geq \sigma_r - 1 + \delta_r = t_r - 1 = \ell - 1$ for a non-leaf MC).

The capacity violation of factor 3 can happen in the case when no facility is opened in $\tau(j')$ for $j' \in \mathcal{C}_S$ and $\psi(j') \in \mathcal{C}_D$ as explained in Lemma (12).

Leaf meta-clusters may have length less than l but they do not have any demand coming onto them from the children meta-cluster, thus capacity violation is bounded as explained in Lemma (12).

(ii) Let j' belongs to a MC G_r such that its demand is not served within G_r. Then, j' must be the root of G_r and its demand is served by facilities in clusters of the parent MC, say G_s. Since the edges in G_s are no costlier than the connecting edge $(j', \psi(j'))$ of G_r and there are at most $\ell - 1$ edges in G_s, the total distance traveled by demand $d_{j'}$ of j' to reach the centers of the clusters in which they are served is bounded by $\ell d_{j'} c(j',\ \psi(j'))$.

5.5 Proof of Lemma 14

Let $j' \in \mathcal{C}'$. Let $\lambda(j')$ be the set of centers j'' such that facilities in $\tau(j'')$ serve the demand of j'. Note that if some facility is opened in $\tau(j')$, then $\lambda(j')$ is $\{j'\}$ itself and if no facility is opened in $\tau(j')$, then $\lambda(j') = \{j'' : \exists i \in \tau(j'')$ such that demand of j' is served by i as per the assignments done in Lemmas (12) and (13)$\}$.

The cost of assigning a part of the demand $d_{j'}$ to a facility opened in $\lambda(j') \cap \mathcal{C}_S$ is bounded differently from the part assigned to facilities in $\lambda(j') \cap \mathcal{C}_D$.

Let $j'' \in \mathcal{C}_S \cap \lambda(j'),\ i \in \tau(j'')$. Then, $c(j'',\ i) \leq c(j'',\ \psi(j'')) \leq c(j',\ \psi(j'))$. Last inequality follows as: either j'' is above j' in the same MC (say G_r) (by Lemma (12.3)) or j'' is in the parent MC (say G_s) of G_r. In the first case, the edge $(j'',\ \psi(j''))$ is either in G_r or is the connecting edge of G_r. The inequality follows as edge costs are non-increasing as we go up the tree. In the latter case, edge $(j'', \psi(j''))$ is either in G_s or it is the connecting edge of G_s: in either case, $c(j'', \psi(j'')) \leq c(j', \psi(j'))$ as the connecting edge of G_s is no costlier than the edges in G_s which are no costlier than the connecting edge of G_r (possibly $c(j', \psi(j'))$) which are no costlier than the edges in G_r. Summing over all $j', j'' \in \mathcal{C}_S$, we see that this cost is bounded by $O(1)LP_{opt}$.

Next, let $j'' \in \mathcal{C}_D \cap \lambda(j'), i \in \mathcal{N}_{j''}$. Further, let g_i be the total demand served by a facility i. Since $g_i \leq 3u$, the cost of transporting $3u$ units of demand from j'' to i is $3u\hat{w}_i c(i, j'')$. Summing it over all $i \in \mathcal{N}_{j''}$, $j'' \in \mathcal{C}_D$, and then over all $j' \in \mathcal{C}'$, we get that the total cost for $\mathcal{C}_D$ is bounded by $O(1)LP_{opt}$.

5.6 Proof of Lemma 17

(i) Let $j_d \in \mathcal{C}_D \cap G_r^1$. Consider the case when $res(j_d) < \epsilon$. The total demand $(\lfloor d_{j_d}/u \rfloor + res(j_d))u \leq (\lfloor d_{j_d}/u \rfloor + \epsilon)u$ of G_r^1 can be distributed to the opened facilities $(\geq \lfloor d_{j_d}/u \rfloor)$ at a loss of factor 2 in capacity as $\lfloor d_{j_d}/u \rfloor \geq 1$.

When $\epsilon \leq res(j_d) < 1$, the demand of G_r^1 is at most $(\lfloor d_{j_d}/u \rfloor + res(j_d) + 1)u \leq (\lfloor d_{j_d}/u \rfloor + 2)u$. The available opening is $\lfloor d_{j_d}/u \rfloor + 1$. Thus, the capacity violation is at most $(\lfloor d_{j_d}/u \rfloor + 2)u/(\lfloor d_{j_d}/u \rfloor + 1)u < 2$ as $\lfloor d_{j_d}/u \rfloor \geq 1$. Hence G_r^1 is self-sufficient.

For $\sigma_r = 0$, (ii) - (vi) hold vacuously. Thus, now onwards we assume that $\sigma_r \geq 1$ (ii) LP_2 opens $\max\{0, \sigma'_r - 1\}$ facilities in G_r^2 where σ'_r is the number of clusters in G_r^2. Constraint (12) ensures that at most one facility is opened in each cluster. Thus, there is at most one cluster in G_r^2 with no facility opened in it and it is a sparse cluster. Next consider G_r^1 with a sparse cluster in it, i.e., $G_r^1 = \{j_d, j_s\}$, it is possible that all the γ_r facilities are opened in $\tau(j_d)$ and no facility is opened in $\tau(j_s)$. Thus, there are at most two clusters with no facility opened in them and these clusters are sparse. (iii) & (iv) Let $j' \in G_r^2$ such that no facility is opened in $\tau(j')$. If $\psi(j') \in G_r^2$, then LP_2 must have opened a facility in $\tau(\psi(j'))$. Demand of j' is assigned to this facility at a loss of maximum 2 factor in capacity. If $\psi(j') \notin G_r^2$ then either G_r^1 is empty or $\psi(j') \in G_r^1$. In the former case j' must be the root of G_r and G_r cannot be the root MC. Clearly, at most u units of demand of G_r remain unassigned within G_r. In the latter case i.e., $\psi(j') \in G_r^1$, then $\psi(j')$ is either j_d or j_s. We will next show that demand of j' will be absorbed in $\tau(j_d) \cup \tau(j_s)$ in the claimed bounds along with claims (v) and (vi) of the lemma.

1. $res(j_d) < \epsilon$, we have $G_r^1 = \{j_d\}$, $\gamma_r = \lfloor d_{j_d}/u \rfloor$, $G_r^2 = G_r \cap \mathcal{C}_S$, $\sigma'_r = \sigma_r$, and $\beta_r = \lfloor d_{j_d}/u \rfloor + \sigma_r - 1$. In this case, $j' = j_s$ and $\psi(j') = j_d$. LP_2 must have opened at least $\lfloor d_{j_d}/u \rfloor \geq 1$ facilities in $\tau(j_d)$ Total demand $(\lfloor d_{j_d}/u \rfloor + res(j_d) + 1))u$ of j_d and j' can be distributed to the facilities opened in $\tau(j_d)$ $(\geq \lfloor d_{j_d}/u \rfloor)$ at a loss of factor $2 + \epsilon$ in capacity, as $res(j_d) < \epsilon$ and $1 \leq \lfloor d_{j_d}/u \rfloor$.
2. $\epsilon \leq res(j_d) < 1$, we have $G_r^1 = \{j_d, j_s\}$, $\gamma_r = \lfloor d_{j_d}/u \rfloor + 1$, $G_r^2 = G_r \cap \mathcal{C}_S \setminus \{j_s\}$, $\sigma'_r = \sigma_r - 1$ and $\beta_r = \lfloor d_{j_d}/u \rfloor + \sigma_r - 1$ if $\sigma_r \geq 2$ and $= \lfloor d_{j_d}/u \rfloor + 1$ if $\sigma_r = 1$. In this case, $\psi(j') = j_s$. In the worst case, no facility is opened in $\tau(j_s)$. LP_2 must have opened at least $\lfloor d_{j_d}/u \rfloor + 1 \geq 2$ facilities in $\tau(j_d) \cup \tau(j_s)$. Total demand $(\lfloor d_{j_d}/u \rfloor + res(j_d) + 1 + 1))u$ of j_d, j_s and j' can be distributed to the facilities opened in $\tau(j_d) \cup \tau(j_s)$ $(\geq \lfloor d_{j_d}/u \rfloor + 1)$ at a loss of factor 2 in capacity, as $\lfloor d_{j_d}/u \rfloor + 1 \geq 2$.

(vii) Clearly, $c(j', j_d) \leq 2c(j', \psi(j'))$. (2) above also handles the case when no facility is opened in a sparse cluster in G_r^1.