CSE 594: Combinatorial and Graph Algorithms — Lecturer: Hung Q. Ngo
SUNY at Buffalo, Spring 2005 — Last update: May 16, 2005

More Approximation Algorithms Based on Linear Programming

We discuss several other problems with approximation algorithms based on linear programming. Most of these problems require a certain level of ingenuity which does not fit perfectly into the frameworks presented in earlier lectures (rounding, randomized rounding, primal-dual).

1 Multicut in trees

In the MULTICUT problem, we are given a graph $G = (V, E)$ with edge capacity function $c : E \to \mathbb{Z}^+$, and m pairs of vertices (s_i, t_i), $i \in [m]$. The pairs are different, but the vertices in different pairs do not have to be distinct. A *multicut* is a set of edges whose removal disconnects s_i and t_i, for all $i \in [m]$. The problem is to find a multicut with minimum capacity.

Exercise 1. Show that MULTICUT is $\mathbf{NP}$-hard even when G is a tree by a reduction from VERTEX COVER.

Throughout this section, we assume G is a tree, so that there is a unique path P_i from s_i to t_i in G. This can be viewed as a SET COVER problem in which an edge e covers all s_i, t_i-paths that contain e. We will show that the algorithm PRIMAL-DUAL WITH REVERSE DELETION gives an approximation ratio 2 for this problem.

The LP-relaxation of the IP for this problem is

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e x_e \\ \text{subject to} \quad & \sum_{e \in P_i} x_e \geq 1 \quad i \in [m], \\ & x_e \geq 0, \quad \forall e \in E. \end{aligned} \tag{1}$$

The dual program is

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} y_i \\ \text{subject to} \quad & \sum_{i:\, e \in P_i} y_i \leq c_e \qquad e \in e, \\ & y_i \geq 0, \quad \forall i \in [m]. \end{aligned} \tag{2}$$

In the context of this problem, we need to be very specific on choosing an uncovered s_k, t_k-path at each iteration. Fix a vertex r of G as the root of the tree. For each pair (s_i, t_i), let $\text{LCA}(s_i, t_i)$ denote the *least common ancestor* of s_i and t_i, which is the vertex at the intersection of the paths from s_i to r and from t_i to r. Note that $\text{LCA}(s_i, t_i)$ could be s_i or t_i. Let $d(s_i, t_i)$ denote the *depth* of $\text{LCA}(s_i, t_i)$, which is the distance between r and $\text{LCA}(s_i, t_i)$. Our primal-dual based algorithm is as follows.

PRIMAL-DUAL FOR MULTICUT IN TREES

$C \leftarrow \emptyset$, $y \leftarrow 0$, $j \leftarrow 0$
while C is not a multicut **do**
  Choose an uncovered path P_k with largest $d(s_k, t_k)$

4: Increase y_k until there is a saturated edge e_j
5: Add e_j into C
6: **end while**
7: $\overline{C} \leftarrow$ REVERSE-DELETE(C)

Theorem 1.1. *The algorithm above gives approximation ratio* 2.

Proof. Let C be an infeasible set of edges. Let A be a minimal augmentation of C. Let P_k be the uncovered path returned by the algorithm in line 3. We only have to show that there are at most 2 edges of P_k on A. (Recall Theorem 7.4 in the previous lecture.)

Let $v = \text{LCA}(s_k, t_k)$. Suppose there are at least 3 edges of A on P_k. These edges are not in C. Without loss of generality, assume there are two edges e_1 and e_2 of A on the part of P_k from s_k to v (otherwise, there are at least two edges on the part from t_k to v). Suppose e_1 is closer to v than e_2. Since A is minimal, e_2 has to be on some P_i to separate s_i and t_i which C does not disconnect. However, $d(s_i, t_i) > d(s_k, t_k)$, implying that C already disconnect s_i and t_i, otherwise P_i would have been chosen instead of P_k. This is a contradiction. □

Note that the integer version of (2) is the MAXIMUM INTEGER MULTI-COMMODITY FLOW problem in trees. The algorithm above implicitly gives a feasible solution for this problem (defined by the y_k). This feasible solution is at least half of the optimal cost of the primal IP, hence it is also a 2 approximation for the multi-commodity flow problem.

Open Problem 1. No non-trivial approximation algorithm is known for the MAXIMUM INTEGER MULTI-COMMODITY FLOW problem on graphs more general than trees. (Trivial ones give factor $\Omega(n)$.)

Exercise 2. Suppose instead of doing REVERSE-DELETE (which deletes edges in the reverse order of which they were added into C), we apply the following procedure: sort edges in the final C by decreasing capacity and remove redundant edges in this order. What factor can you prove for the modified algorithm?

Exercise 3. Give a polynomial time algorithm to compute a maximum integer multi-commodity flow on trees with unit edge-capacities. You can use as a subroutine a maximum matching algorithm. (**Hint**: dynamic programming.)

2 Metric uncapacitated facility location

FACILITY LOCATION is a fundamental optimization problem appearing in various context. In the uncapacitated version of the problem, we are given a complete bipartite graph $G = (F, C; E)$ where F represents a set of "facilities" and C a set of "cities." The cost of opening facility i is f_i, and the cost of assigning city j to facility i is c_{ij}. The problem is to find a subset $I \subseteq F$ of facilities to be open and an assignment $a : C \to I$ assigning every city j to some facility $a(j)$ to minimize the cost function

$$\sum_{i \in I} f_i + \sum_{j \in C} c_{a(j),j}.$$

In the metric version of the problem, the cost c_{ij} satisfies the triangle inequality.

Designate a variable x_i indicating if facility i is open and y_{ij} indicating if city j is assigned to facility

i, we get the following integer program:

$$\begin{array}{lll} \min & \sum_{i\in F} f_i x_i + \sum_{i\in F, j\in C} c_{ij} y_{ij} & \\ \text{subject to} & \sum_{i\in F} y_{ij} \geq 1 & j \in C, \\ & x_i - y_{ij} \geq 0 & i \in F,\ j \in C, \\ & x_i, y_{ij} \in \{0,1\}, & i \in F,\ j \in C. \end{array} \tag{3}$$

Relaxing this integer program gives the following linear program

$$\begin{array}{lll} \min & \sum_{i\in F} f_i x_i + \sum_{i\in F, j\in C} c_{ij} y_{ij} & \\ \text{subject to} & \sum_{i\in F} y_{ij} \geq 1 & j \in C, \\ & x_i - y_{ij} \geq 0 & i \in F,\ j \in C, \\ & x_i, y_{ij} \geq 0, & i \in F,\ j \in C. \end{array} \tag{4}$$

The dual linear program is

$$\begin{array}{lll} \max & \sum_{j\in C} s_j & \\ \text{subject to} & \sum_{j\in C} t_{ij} \leq f_i & i \in F, \\ & s_j - t_{ij} \leq c_{ij} & i \in F,\ j \in C, \\ & s_j, t_{ij} \geq 0, & i \in F,\ j \in C. \end{array} \tag{5}$$

This primal-dual pair does not fit nicely into the covering primal-dual approximation framework we discussed. In particular, there are negative coefficients. The general idea of applying the primal-dual method to this problem is still to find some sort of “maximal” dual-feasible solution, and then set to 1 the primal variables corresponding to saturated primal constraints.

PRIMAL-DUAL FOR METRIC UNCAPACITATED FACILITY LOCATION

Phase 1.

$O \leftarrow \emptyset$; // set of temporarily open facilities
$J \leftarrow \emptyset$; // set of connected cities thus far
$\mathbf{s} \leftarrow 0$; $\mathbf{t} \leftarrow 0$
while $J \neq C$ **do**
  Increase uniformly all s_j, $j \in C - J$
  After a while, if for some edge ij, we reach $s_j - t_{ij} = c_{ij}$, then increase uniformly t_{ij} also.
  // Edges with $s_j - t_{ij} = c_{ij}$ are called “tight” Edges with $t_{ij} > 0$ are called “special”
  As soon as an edge ij becomes tight, if $i \in O$ then add j into J and declare i the “connection witness” for j
  After a while, there is some i such that $f_i = \sum_{j\in C} t_{ij}$.
  for each such i in any order **do**
    $O \leftarrow O \cup \{i\}$
    for each tight edge ij with $j \notin J$ **do**
      $J \leftarrow J \cup \{j\}$ // i is called a “connection witness” for j
    end for
  end for

end while

Phase 2.

Let $H = (O, C)$ be the bipartite graph containing only special edges
Let I be a maximal subset of O such that there is no path of length 2 in H between any two vertices in I
for each $j \in C$ **do**
  if $\exists i \in I$ such that ij is special **then**
    $a(j) \leftarrow i$ // call j "directly connected" to i
  else
    Let i be the connection witness for j
    if $i \in I$ **then**
      $a(j) \leftarrow i$ call j "directly connected" to i // note that ij is tight but not special
    else
      There must be some $i' \in I$ within H-distance 2 from i
      $a(j) \leftarrow i'$ call j "indirectly connected" to i'
    end if
  end if
end for

We shall use $(\bar{\mathbf{s}}, \bar{\mathbf{t}})$ to denote the returned dual-feasible solution $(\mathbf{s}, \mathbf{t})$.

Theorem 2.1. *The algorithm above gives approximation ratio* 3*.*

Proof. The idea is to compare the cost of the approximated solution

$$\sum_{i \in I} f_i + \sum_{j \in C} c_{a(j),j}$$

to the cost of the dual-feasible solution $(\bar{s}, \bar{t})$, which is $\sum_{j \in C} \bar{s}_j$.

We shall break each $\bar{s}_j$ into two parts, and write $\bar{s}_j = \bar{s}_j^f + \bar{s}_j^c$ in the following way. If j is directly connected to $i = a(j)$, then set $\bar{s}_j^f = \bar{t}_{ij}$ and $\bar{s}_j^c = c_{ij}$. If j is indirectly connected to $i = a(j)$, then set $\bar{s}_j^f = 0$ and $\bar{s}_j^c = \bar{s}_j$. Intuitively, the term $\bar{s}_j^f$ is the contribution of j into opening facility i, and the term $\bar{s}_j^c$ is the contribution of j to the cost of having edge ij.

Firstly, if $i \in I$ and ij is special, then j is directly connected to i. Consequently

$$\sum_{i \in I} f_i = \sum_{i \in I} \sum_{j : ij \text{ special}} \bar{t}_{ij} = \sum_{j \in C} \bar{s}_j^f.$$

Secondly, we claim that $c_{ij} \leq 3\bar{s}_j^c$, where $i = a(j)$. If j is directly connected to i, then $c_{ij} = \bar{s}_j^c$ by definition. When j is indirectly connected to i, there is an $i' \in I$, $j' \in C$ such that ij' and $i'j'$ are special, and that i' is a connection witness for j. By the triangle inequality, it is sufficient to prove that all three of $c_{i'j}$, $c_{i'j'}$, and $c_{ij'}$ are at most $\bar{s}_j$.

Since i' is a connection witness for j, the edge $i'j$ is tight, implying $c_{i'j} \leq \bar{s}_j$. If we can show that $s_{j'} \leq s_j$, then the other two inequalities follow. Since i' is a connection witness for j, s_j got increased until right at or after the time i' became temporarily open. Since both $i'j'$ and ij' are special, $s_{j'}$ could not have gotten increased after the time i' became temporarily open. We thus have $s_{j'} \leq s_j$.

Consequently,

$$\sum_{i \in I} f_i + \sum_{j \in C} c_{a(j),j} \leq \sum_{j \in C} \left(\bar{s}_j^f + 3\bar{s}_j^c \right) \leq 3\text{OPT}.$$

□

Exercise 4. The vector $\bar{\mathbf{s}}$ found by the algorithm above is maximal in the sense that, if we increase any $\bar{s}_j$ and keep other $\bar{s}_j$ the same, then there is no way to adjust the $\bar{t}_{ij}$ so that $(\bar{\mathbf{s}}, \bar{\mathbf{t}})$ is still dual-feasible. Is every maximal solution $\bar{\mathbf{s}}$ within 3 times the optimal solution to the dual linear program?

Hint: consider n facilities with opening cost of 1 each, n cities connected to distinct facilities with cost ϵ each. In addition, there is another city that is connected to each facility with an edge of cost 1.

Exercise 5. Suppose the cost of connecting city i to facility j is c_{ij}^2, where the costs c_{ij} still satisfy the triangle inequality (but their squares may not). Show that our algorithm gives performance ratio 9.

Exercise 6. Suppose we slightly change the problem in the following way. Each city j has a demand d_j. The cost of connecting j to an open facility i is now $c_{ij}d_j$. (Previously, all d_j are 1.) Modify our algorithm to get a 3-approximation for this problem. (**Hint**: raise s_j at rate d_j.)

3 Metric k-median

The k-MEDIAN problem is very similar to the FACILITY LOCATION problem. The difference is that there is no cost for opening facilities. On the other hand, there is an upper bound of k on the number of open facilities.

Keeping x_i and y_{ij} as in the previous section, we can obtain an integer program for the k-MEDIAN problem. Its LP-relaxation is as follows.

$$\begin{array}{lll} \min & \sum_{i\in F, j\in C} c_{ij}y_{ij} & \\ \text{subject to} & \sum_{i\in F} y_{ij} \geq 1 & j \in C, \\ & x_i - y_{ij} \geq 0 & i \in F,\ j \in C, \\ & \sum_{i\in F}(-x_i) \geq -k & \\ & x_i, y_{ij} \geq 0, & i \in F,\ j \in C. \end{array} \tag{6}$$

The dual linear program is

$$\begin{array}{lll} \max & \sum_{j\in F} s_j - ku & \\ \text{subject to} & \sum_{j\in C} t_{ij} \leq u & i \in F, \\ & s_j - t_{ij} \leq c_{ij} & i \in F,\ j \in C, \\ & s_j, t_{ij}, u \geq 0, & i \in F,\ j \in C. \end{array} \tag{7}$$

This primal-dual pair looks strikingly similar to the primal-dual pair of the FACILITY LOCATION problem. In fact, if we assign a cost of u to each facility in the FACILITY LOCATION problem and solve for the primal optimal solution $(\mathbf{x}, \mathbf{y})$ and dual optimal solution $(\mathbf{s}, \mathbf{t})$, then by strong duality

$$\sum_{i\in F} ux_i + \sum_{i\in F, j\in C} c_{ij}y_{ij} = \sum_{j\in F} s_j.$$

Consequently, if there is a value of u such that the primal optimal solution $(\mathbf{x}, \mathbf{y})$ opens exactly k facilities (fractionally), i.e. $\sum_{i\in F} x_i = k$, then it is clear that $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{s}, \mathbf{t}, u)$ are optimal solutions to the primal-dual pair of the k-MEDIAN problem.

On the same line of thought, suppose we can find a value of u for which the approximation algorithm for FACILITY LOCATION returns an integral solution $(\mathbf{x}, \mathbf{y})$ and a dual-feasible solution $(\mathbf{s}, \mathbf{t})$ such that exactly k facilities are open, then

$$\begin{aligned} 3\sum_{i\in F} ux_i + \sum_{i\in F, j\in C} c_{ij}y_{ij} &= 3\sum_{i\in I} f_i + \sum_{j\in C} c_{a(j),j} \\ &\le 3\sum_{j\in C}(s_j^f + s_j^c) \\ &= 3\sum_{j\in C} s_j. \end{aligned}$$

This implies

$$\sum_{i\in F, j\in C} c_{ij}y_{ij} = \sum_{j\in C} c_{a(j),j} \le 3\left(\sum_{j\in C} s_j - ku\right),$$

and we would have gotten a 3-approximation algorithm for the k-MEDIAN problem. Unfortunately, it is an open problem to find a u so that this happens. We will take a different path.

Let n_c be the number of cities, n_f number of facilities, $n = n_c + n_f$, and $m = n_c n_f$ the number of edges in the graph.

In the algorithm for FACILITY LOCATION, the larger u is the fewer number of facilities will be opened. (More edges will become tight before the cost u is reached.) When $u = 0$, all facilities will be opened. When $u = n_c c_{\max}$, where $c_{\max}$ is the maximum c_{ij} and n is the number of cities, only one facility will be opened, because all edges are tight when this cost is reached. Assuming we break ties canonically, it is easy to see that the number of opened facilities is inversely proportional to u.

Apply binary search on the interval $[0, n_c c_{\max}]$ to find two values $u_1 < u_2$ such that the corresponding number of opened facilities k_1, k_2 satisfy $k_1 > k > k_2$ and that $u_2 - u_1 \le c_{\min}/(12n_f^2)$, where $c_{\min}$ is the minimum value of c_{ij}. (If we can find a value of u for which the number of opened facilities is k, then we are done.) Let the corresponding integral primal solutions be $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)})$ and $(\mathbf{x}^{(2)}, \mathbf{y}^{(2)})$, and the corresponding (fractional) dual solutions be $(\mathbf{s}^{(1)}, \mathbf{t}^{(1)})$ and $(\mathbf{s}^{(2)}, \mathbf{t}^{(2)})$, respectively.

First, the idea is to get a convex combination

$$(\mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}) + \beta(\mathbf{x}^{(2)}, \mathbf{y}^{(2)})$$

such that $(\mathbf{x}, \mathbf{y})$ opens k facilities fractionally. This means $\alpha k_1 + \beta k_2 = k$ and $\alpha + \beta = 1$, which implies $\alpha = \frac{k-k_2}{k_1-k_2}$, and $\beta = \frac{k_1-k}{k_1-k_2}$. Let

$$(\mathbf{s}, \mathbf{t}) = \alpha(\mathbf{s}^{(1)}, \mathbf{t}^{(1)}) + \beta(\mathbf{s}^{(2)}, \mathbf{t}^{(2)}).$$

Note that $(\mathbf{s}, \mathbf{t}, u_2)$ is a feasible solution to the dual program (7).

Let us estimate how good this fractional combination is. We have

$$\sum_{i\in F, j\in C} c_{ij}y_{ij}^{(1)} \le 3\left(\sum_{j\in C} s_j^{(1)} - k_1u_1\right) \tag{8}$$

$$\sum_{i\in F, j\in C} c_{ij}y_{ij}^{(2)} \le 3\left(\sum_{j\in C} s_j^{(2)} - k_2u_2\right). \tag{9}$$

We want to turn u_1 in the first inequality into u_2 with a small increase in the factor 3, and then take the convex combination to estimate the cost of $(\mathbf{x}, \mathbf{y})$. Using the facts that $u_2 - u_1 \le \frac{c_{\min}}{12n_f^2}$, $k_1 \le n_f$, and

$c_{\min} \le \sum_{i\in F, j\in C} c_{ij} y_{ij}^{(1)}$, we get

$$\sum_{i\in F, j\in C} c_{ij} y_{ij}^{(1)} \le (3 + 1/n_f)\left(\sum_{j\in C} s_j^{(1)} - k_1 u_2\right). \tag{10}$$

Now, an α, β convex combination of (10) and (9) gives

$$\sum_{i\in F, j\in C} c_{ij} y_{ij} \le (3 + 1/n_f)\left(\sum_{j\in C} s_j - k u_2\right).$$

Thus, the fractional solution $(\mathbf{x}, \mathbf{y})$ is within $(3 + 1/n_f)$ of the optimal. To turn the fractional solution into an integral solution, we apply randomized rounding.

Let I_1 and I_2 be the set of facilities returned by the algorithm corresponding to u_1 and u_2 respectively. We know $|I_1| = k_1$ and $|I_2| = k_2$, and that $k_1 > k > k_2$. The fractions α and β indicate how much the solution should be leaning towards I_1 and I_2. Hence, it is natural to use them as rounding probability. The trouble is that I_2 does not have enough elements, while I_1 has too many elements.

We resolve this problem in the following way. For each i in I_2, let $\nu(i)$ be a facility in I_1 nearest to i. Let I_ν be the set of these $\nu(i)$. Clearly $|I_\nu| \le k_2 < k_1$. We arbitrarily pad I_ν with elements from I_1 until $|I_\nu| = k_2$. Our rounding procedure goes as follows.

- Open all facilities in I_2 with probability β and all facilities in I_ν with probability α.
- Pick uniformly a subset I_3 of $I_1 - I_\nu$ of size $|I_3| = k - k_2$ and open all facilities in I_3. Note that each element in $I_1 - I_\nu$ has a probability of

  $$\frac{\binom{k_1-k_2-1}{k-k_2-1}}{\binom{k_1-k_2}{k-k_2}} = \frac{k-k_2}{k_1-k_2} = \alpha$$

  of being chosen.
- Return the set I of k opened facilities.

The next thing to do is to assign cities to these opened facilities. Consider any city j. Let i_1 and i_2 be the facilities that j was connected to in the solutions I_1 and I_2. In the first case, suppose $i_1 \in I_\nu$. Since either i_1 or i_2 is open, we set $a(j)$ to be the open facility. In the second case, suppose $i_1 \notin I_\nu$, in which case we connect j to i_1 if it is open (i.e. $i_1 \in I_3$), otherwise to i_2 if it is open. If both i_1 and i_2 are not open, we connect j to $i_3 = \nu(i_2)$.

We estimate the expected cost of connecting j to $a(j)$. In the first case ($i_1 \in I_\nu$),

$$\mathrm{E}[c_{a(j),j}] = \alpha c_{i_1 j} + \beta c_{i_2 j}.$$

In the second case when $i_1 \not{i}n I_\nu$, there is a probability of α that i_1 is in I_3, a probability of $(1-\alpha)\beta = \beta^2$ that $i_1 \notin I_3$ but i_2 is open, and a probability of $(1-\alpha)(1-\beta) = \alpha\beta$ that j will be connected to i_3. Thus, in this case

$$\mathrm{E}[c_{a(j),j}] = \alpha c_{i_1 j} + \beta^2 c_{i_2 j} + \alpha\beta c_{i_3 j}.$$

By the triangle inequality, we have

$$c_{i_3 j} \le c_{i_3 i_2} + c_{i_2 j} \le c_{i_1 i_2} + c_{i_2 j} \le c_{i_1 j} + 2 c_{i_2 j}.$$

Consequently,

$$\mathrm{E}[c_{a(j),j}] \le \alpha(1+\beta) c_{i_1 j} + \beta(1+\alpha) c_{i_2 j} \le (1 + \max\{\alpha, \beta\})[\alpha c_{i_1 j} + \beta c_{i_2 j}].$$

We have just shown the following theorem.

Theorem 3.1. *The above rounding procedure gives expected cost at most* $(1 + \max\{\alpha, \beta\})$ *the cost of* $(\mathbf{x}, \mathbf{y})$.

Thus, in total the rounded solution is of cost at most $(3+1/n_f)(1+\max\{\alpha,\beta\})$ of the optimal. Since $\max\{\alpha, \beta\}$ is at most $n_f/(n_f+1)$, and $(3+1/n_f)(1+n_f/(n_f+1)) \le 6$, we obtain a 6-approximation.

The algorithm can be derandomized with the method of conditional expectation. Note that the expectations $\mathrm{E}[c_{a(j),j}]$ can be calculated explicitly and efficiently. To derandomize this algorithm, we compute the expectations of the final cost given that we open I_2 or I_ν. We then follow the smaller expectation (in the α or β weighted sense). To compute which I_3 to open, we can compute the conditional expectations of the cost given that elements $i_1, \dots, i_{k_2-k_1}$ of $I_1 - I_2$ are in I_3, then repeat this process.

Exercise 7 (Vazirani's book - Exercise 25.3)**.** Use Lagrangian relaxation technique to give a constant factor approximation algorithm for the following common generalization of the FACILITY LOCATION and k-MEDIAN problems. Consider the UNCAPACITATED FACILITY LOCATION problem with the additional constraint that at most k facilities can be opened.

Exercise 8 (Jain and Vazirani [10])**.** Consider the l_2^2-CLUSTERING problem. Given a set of n points $S = \{v_1, \dots, v_n\}$ in $\mathbb{R}^d$ and a positive integer k, the problem is to find a minimum cost k-clustering, i.e., to find k points, called *centers*, $f_1, \dots, f_k \in \mathbb{R}^d$, so as to minimize the sum of squares of distances from each point v_i to its closest center. This naturally defines a partitioning of the n points into k clusters. Give a constant factor approximation algorithm for this problem.

(**Hint**: first show that restricting the centers to a subset of S increases the cost of the optimization solution by a factor of at most 2.)

Historical Notes

The 2-approximation for multicut in trees was due to Garg, Vazirani, and Yannakakis [6]. Recent works on integer multi-commodity flow can be found in [3–5, 7, 14]. For an example of multi-commodity flow in networking, see [12].

For the UNCAPACITATED FACILITY LOCATION problem, Hochbaum [8] obtained ratio $O(\lg n)$, Shmoys, Tardos, and Aardal [13] got 3.16 ratio with an LP-rounding based algorithm. The 3-approximation algorithm we described was due to Jain and Vazirani [10]. Jain, Mahdian and Saberi [9] reduce the ratio further to 1.61 with a greedy algorithm analyzed by the dual-fitting method.

For the METRIC k-MEDIAN problem, Bartal [] gave the first algorithm which achieved approximation ratio $O(\lg n \lg\lg n)$. Charikar, Guha, Tardos, and Shmoys [2] achieved $6\frac{2}{3}$ using ideas from Lin and Vitter [11]. The 6-approximation algorithm we described was due to Jain and Vazirani [10]. Arya, Garg, Khandekar, Meyerson, Munagala, and Pandit [1] achieved ratio $(3 + 2/p)$ with running time $O(n^p)$, for any p using the local search method. Jain, Mahdian, and Saberi [9] gave a hardness ratio of $1 + 2/e$ for approximating this problem.

References

[1] V. ARYA, N. GARG, R. KHANDEKAR, A. MEYERSON, K. MUNAGALA, AND V. PANDIT, *Local search heuristics for k-median and facility location problems*, SIAM J. Comput., 33 (2004), pp. 544–562 (electronic).

[2] M. CHARIKAR, S. GUHA, É. TARDOS, AND D. B. SHMOYS, *A constant-factor approximation algorithm for the k-median problem*, J. Comput. System Sci., 65 (2002), pp. 129–149. Special issue on STOC, 1999 (Atlanta, GA).

[3] C. CHEKURI AND S. KHANNA, *Edge disjoint paths revisited*, in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), New York, 2003, ACM, pp. 628–637.

[4] C. CHEKURI, S. KHANNA, AND F. B. SHEPHERD, *The all-or-nothing multicommodity flow problem*, in STOC ’04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, New York, NY, USA, 2004, ACM Press, pp. 156–165.

[5] C. CHEKURI, S. KHANNA, AND F. B. SHEPHERD, *Edge-disjoint paths in undirected planar graphs*, in FOCS ’04 (Rome, Italy, ACM, New York, 2004, pp. ??–?? (electronic).

[6] N. GARG, V. V. VAZIRANI, AND M. YANNAKAKIS, *Primal-dual approximation algorithms for integral flow and multicut in trees*, Algorithmica, 18 (1997), pp. 3–20.

[7] V. GURUSWAMI, S. KHANNA, R. RAJARAMAN, B. SHEPHERD, AND M. YANNAKAKIS, *Near-optimal hardness results and approximation algorithms for edge-disjoint paths and related problems*, J. Comput. System Sci., 67 (2003), pp. 473–496.

[8] D. S. HOCHBAUM, *Heuristics for the fixed cost median problem*, Math. Programming, 22 (1982), pp. 148–162.

[9] K. JAIN, M. MAHDIAN, AND A. SABERI, *A new greedy approach for facility location problems*, in STOC ’02: Proceedings of the thiry-fourth annual ACM Symposium on Theory of Computing, ACM Press, 2002, pp. 731–740.

[10] K. JAIN AND V. V. VAZIRANI, *Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation*, J. ACM, 48 (2001), pp. 274–296.

[11] J.-H. LIN AND J. S. VITTER, *Approximation algorithms for geometric median problems*, Inform. Process. Lett., 44 (1992), pp. 245–249.

[12] A. E. OZDAGLAR AND D. P. BERTSEKAS, *Optimal solution of integer multicommodity flow problems with application in optical networks*, in Frontiers in global optimization, vol. 74 of Nonconvex Optim. Appl., Kluwer Acad. Publ., Boston, MA, 2004, pp. 411–435.

[13] D. B. SHMOYS, É. TARDOS, AND K. AARDAL, *Approximation algorithms for facility location problems (extended abstract)*, in STOC ’97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing, ACM Press, 1997, pp. 265–274.

[14] K. VARADARAJAN AND G. VENKATARAMAN, *Graph decomposition and a greedy algorithm for edge-disjoint paths*, in Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms (SODA’04), Philadelphia, PA, USA, 2004, Society for Industrial and Applied Mathematics, pp. 379–380.