COL870
Project Report

# LOCAL SEARCH HEURISTICS FOR k-MEDIAN AND FACILITY LOCATION PROBLEMS

*VIJAY ARYA, NAVEEN GARG, ROHIT KHANDEKAR, ADAM MEYERSON, KAMESH MUNAGALA AND VINAYAKA PANDIT*

Akanksha Jain
2010CS50273

## INTRODUCTION

The paper discusses local search techniques for solving facility location problem( including its special case **k-median** ). The problem can be stated as:
We need to select and open at max **k** facilities out of a given set of possible facilities to cater for a given set of clients. This involves **service cost** and might involve **opening cost** for each facility and for each client with respect to its assigned facility( commonly distance ). Objective is to serve all the clients and minimise the total cost involved.
**Locality gap:** Ratio of the maximum of local optimum solution to the global optimum solution.

There are three different problems considered in this paper:

1. K-medians problem:- Simply minimising the sum of absolute distances of clients and assigned facilities. Here the facility opening cost is not involved.
    a. Single-Swap local search: Locality gap of 5 has been proved.
    b. Multiple-swap: Locality gap of (3+2/p) where **p** is the number of swaps in one step ( tight bound ).
2. UFL:- Uncapacitated Facility Location. The total cost includes opening cost for facilities, but there is no upper bound on number of clients served by each facility. Locality gap of 3 has been achieved.
3. CFL:- Capacitated Facility Location. There is upper bound on the number of clients each facility can serve, and facility opening and copying cost is involved. The algorithm has a locality gap of at most 4 .

**NOTATION AND DEFINITIONS**

We are given two sets: F, set of facilities, and C, the set of clients. $C_{ij}$ may be the cost of serving client $i \in C$ by a facility $j \in F$, which is the distance here.

The objective is, given integer k,

1. Identify a set $S \subseteq F$ of at most k facilities, minimise the cost.
2. For each $i \in F$, we are given a opening cost $f_i \geq 0$ further.
3. for each $i \in F$, an integer capacity $u_i > 0$, which is the maximum number of clients that a single copy of the facility i can serve.

Let B(S) be the neighbourhood of set S. $B: S \rightarrow 2^S$

**ALGORITHM**

The proof of the locality gap proceeds by considering a suitable, polynomially large (in the input size) subset $Q \subseteq B(S)$.

**Algorithm** `Local Search.`

1. $S \leftarrow$ an arbitrary feasible solution in $\mathcal{S}$.
2. While $\exists\, S' \in \mathcal{B}(S)$ such that $cost(S') < cost(S)$,
   do $S \leftarrow S'$.
3. return $S$.

FIG. 1. *A generic local search algorithm.*

The step 2 is modified to limit the number of iterations to polynomial number of steps.

2. While $\exists S \in B(S)$ such that $cost(S) \leq (1 - e/Q)cost(S)$,
   do $S \leftarrow S$

Here $e>0$ is a constant and $Q=|Q|$.

Now the number of steps in the iteration is at most $\log(cost(S_0)/cost(O))/\log(1/(1-e/Q))$.

All three terms are polynomial in size, hence, running time is polynomial.

Given a solution A, let $A_j$ denote the service cost of client j, which is the distance between j and the facility in A which serves it. For every facility $a \in A$, we use $N_A(a)$ to denote the set of clients that a serves.

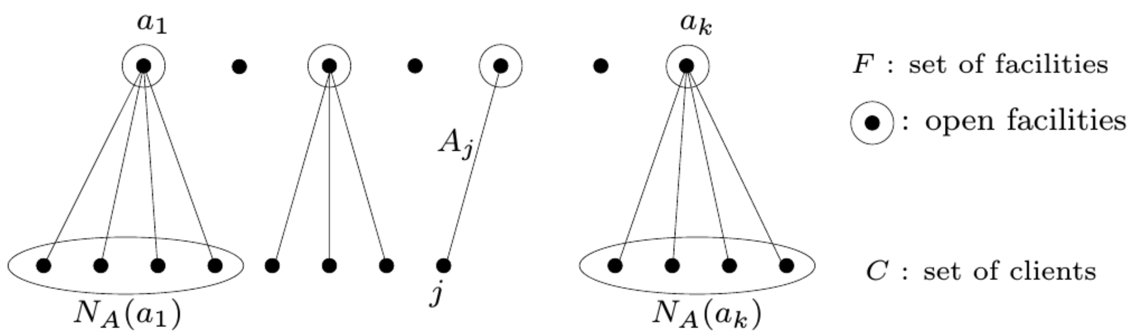


FIG. 2. *Illustration of neighborhood and service costs.*

1. **k-median problem:**

   a. **Local search with single swap**

A swap is effected by closing a facility $s \in S$ and opening a facility $s' \notin S$ and is denoted by <s,s’>, hence $B(S) = \{ S - \{s\} + \{s'\} \mid s \in S \}$

Locality gap of 5 has been proved by using:

Optimality condition of solution S: ( here O is the global optimum solution)

$$cost(S - s + o) \geq cost(S) \text{ for all } s \in S,\ o \in O.$$

Introducing the concept of ‘*capture*’ of a $o \in O$ by $s \in S$, if s has more than half the clients of o. Thus, a facility $s \in S$ captures a facility $o \in O$ if s serves more than half the clients served by o, that is,

$$|N_O(o) \cap N_S(s)| > (|N_O(o)|)/2$$

If s captures any facility in O, it is called a **bad facility**, otherwise it is a **good facility**

Partitioning the set No(O) into sets intersected by Ns(S). Defining a function **$\pi$:No(o)→No(o),** such that if some facility s does not capture facility o, then any point p∈No(o)∩Ns(s) will have its mapping $\pi$:p→q such that (q)∩(No(o)∩Ns(s)) = ∅

Now, we consider the swap procedure in such a way:

1. Each o∈O is considered in exactly one swap.
2. A facility s∈S which captures more than one facility in O is not considered in any swap
3. Each good facility s∈S is considered in at most two swaps.
4. If swap o is considered, then facility s does not capture any facility o'!=o

A bipartite graph is drawn with S and O along the notion of *'capture',* to visualise the swaps considered.

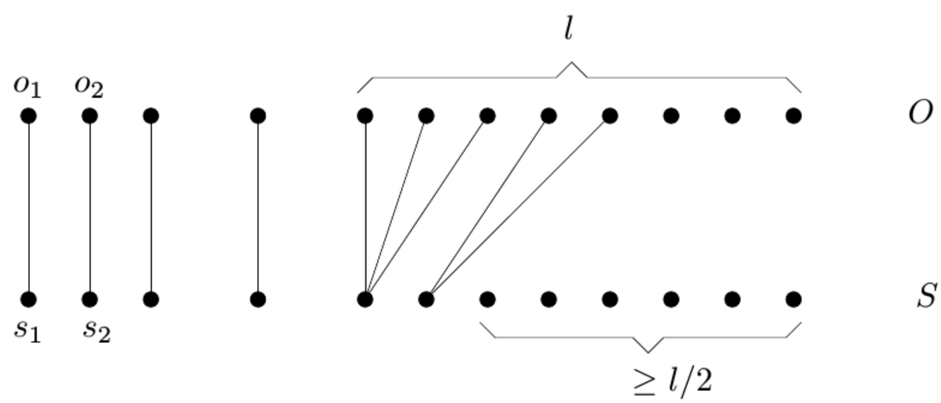


FIG. 5. *Capture graph H = (S, O, E).*

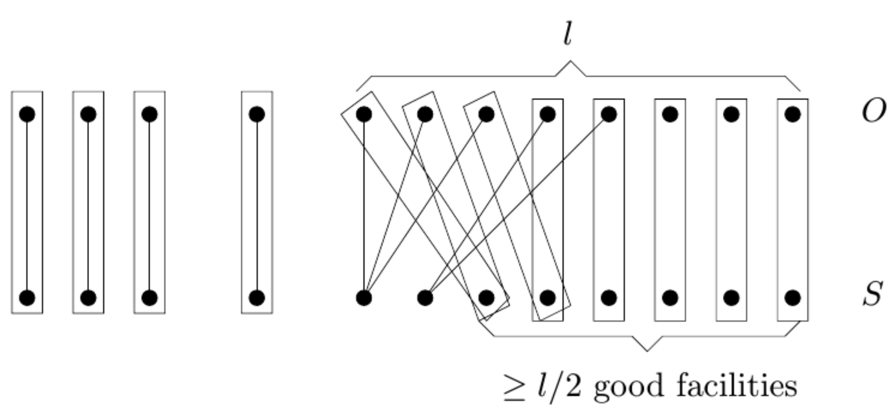


FIG. 6. *k swaps considered in the analysis.*

Now, the swaps considered for good and bad facilities are counted, triangle inequality is used and the locality gap bound is proven.

**b. Multiswap K median**

We divide the set S and O into subsets by the following algorithm:

**procedure** Partition;
$i = 0$
**while** $\exists$ a bad facility in $S$ **do**
**1.** $i = i + 1$ {iteration $i$}
**2.** $A_i \leftarrow \{b\}$ where $b \in S$ is any bad facility
**3.** $B_i \leftarrow capture(A_i)$
**4. while** $|A_i| \neq |B_i|$ **do**
**4.1.** $A_i \leftarrow A_i \cup \{g\}$ where $g \in S \setminus A_i$ is any good facility
**4.2.** $B_i \leftarrow capture(A_i)$
**5.** $S \leftarrow S \setminus A_i$
$O \leftarrow O \setminus B_i$
$A_r \leftarrow S$
$B_r \leftarrow O$
**end.**

FIG. 8. *A procedure to define the partitions.*

Now, we give weights to the swapping of each Ai and Bi, and including the optimality condition for S and modification of the previously defined function $\pi$:NO(o)→NO(o), and with a similar analysis like the single-swap problem, the locality gap of 3+2/p is proved.
Tightness of this bound has been proven with an example.

**2. Uncapacitated facility location problem**

The opening cost of each facility **$f_i$** is included in the cost, which has to minimised, the analysis is similar to the previous problems, by updating the swapping procedure according to the position of the points, the locality gap of 3 is proved.
An example is given to prove its tight bound.

**3. Capacitated facility location problem**

Here we limit the number of clients a facility can serve. We can specify the amount of savings generated by closing a facility s, since a copy of facility s can serve only $u_s$ number of clients, this leads us to define a knapsack problem on a profit generated due to opening l copies of a facility s, for which we have to find the maximum subset T of open facilities which can be deleted.

Further, we use the local optimality criterion, and visualise a bipartite flow graph from vertices of set S on one side, and vertices of set O on another, with the flow amount of $N_s(S)$, using triangle inequality and flow parameters, a locality gap of at most 4 is proven.

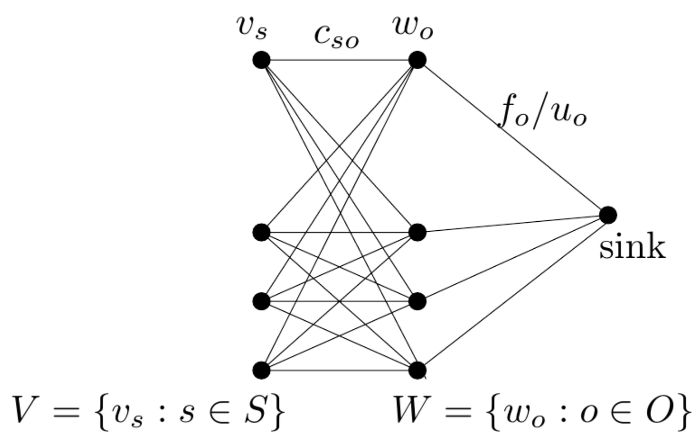


FIG. 14. *The flow graph.*

*All images credit: http://www.cse.iitd.ernet.in/~pandit/lsearchSICOMP.pdf*