**k-Median problem**

- **F**: facilities
- **D**: clients
- **integer** \( k \)

**Feasible solution**: \( k \) facilities \( S \) from \( F \), \( \sigma : D \rightarrow S \)

**Goal**: minimize \( \text{cost}(S, \sigma) = \sum_{j \in D} c_{\sigma(j)}, j \)
Local Search for k-Median

Start with $S_0$ any set of $k$ facilities

Local search move: *swap a facility*

$$S' = S - l + i \text{ where } l \in S \text{ and } i \notin S$$

In each iteration pick best swap and STOP if no improvement possible

(can use approximate stopping condition)
Local Search for k-Median

**Theorem:** Local optima is 5-approximate

Improved local search: swap p facilities at a time

\[ S' = S - A + B \]

where \( |A| = |B| = p \) and \( A \subset S \), \( B \cap S = \emptyset \)

Running time depends on \( n^p \)

**Theorem:** Local optima is \((3+2/p)\) approximate
Analysis for 5-approximation

Clever and perhaps not so intuitive argument

Let $S$ be a local optimum and $O$ be an optimum solution

implies for every $l \in S$ and $i \in O$

$\text{cost}(S-l+i) \geq \text{cost}(S)$

(we ignore the approximate stopping condition in this analysis)
Set up $k$ swap pairs carefully chosen

each facility $i \in O$ is in exactly \textit{one} of the pairs

each facility $l \in S$ is at most \textit{two} of the pairs

Let $\sigma$ and $\omega$ be the assignment functions of clients to facilities for $S$ and $O$
Analysis

A facility $l \in S$ is said to *capture* a facility $i \in O$ iff at least half the clients of $i$ are clients of $l$

in other words $\left| \omega^{-1}(i) \cap \sigma^{-1}(l) \right| \geq \left| \omega^{-1}(i) \right|/2$

Intuitively $l$ is close to $i$

Note: a facility in $O$ is captured by at most one facility in $S$
but a facility in $S$ can capture multiple facilities in $O$
Analysis

Set up a bipartite graph $H=(S,O)$ as follows:
There is an edge $(l, i)$ in $H$ iff $l$ captures $i$

Now we define the $k$ swap pairs:
- Remove all facilities in $S$ with degree $>1$ in $H$.
- If $(l, i)$ is an edge in $H$ and degree of $l$ is 1 then add $(l, i)$ to the swap pair and remove $l, i$.
- Pair each remaining facility $i'$ in $O$ with some arbitrary remaining facility in $S$ s.t. no facility in $S$ is paired more than twice (check that this is possible by simple counting).
Analysis

The final ingredient of the set up is a permutation $\pi$ on the clients that satisfies some important properties

$\pi: \mathcal{D} \rightarrow \mathcal{D}$ permutation implies 1-1 onto function

Properties:
- $j$ and $\pi(j)$ are assigned to the *same* facility in $\mathcal{O}$, that is $\omega(j) = \omega(\pi(j))$
- $j$ and $\pi(j)$ are assigned to *different* facilities in $\mathcal{S}$ unless $\omega(j)$ is captured in which case we don’t care
\[ \omega(j) = \omega(\pi(j)) \]

Diagram:
- \( \sigma(j) \)
- \( \omega(j) \)
- \( \sigma(\pi(j)) \)
- \( j \)
- \( \pi(j) \)
Analysis

Can show that $\pi$ satisfying properties exists

Consider $i \in O$ s.t $i$ is not captured. Note that we can restrict attention to clients in $\omega^{-1}(i)$ and define $\pi$ on them.

For each $j \in \omega^{-1}(i)$ we need to ensure that $\pi(j)$ has the property that $\sigma(j) \neq \sigma(\pi(j))$
Analysis

For each $j \in \omega^{-1}(i)$ we need to ensure that $\pi(j)$ has the property that $\sigma(j) \neq \sigma(\pi(j))$

Wlog assume clients in $\omega^{-1}(i)$ are numbered 0,1,2,...,h-1 where $h = | \omega^{-1}(i) |$ and further that clients with same $\sigma$ are consecutively numbered.

set $\pi(j) = j + \lceil h/2 \rceil \mod h$

Check that this satisfies the desired property (here is where the fact that $i$ is not captured is used)
Analysis

Now we are ready to prove the 5-approximation

foreach swap pair \((l, i)\),
\[\text{cost}(S - l + i) - \text{cost}(S) \geq 0\]

lower bound \[\sum_{(l, i)} (\text{cost}(S - l + i) - \text{cost}(S))\]
by
5 \(\text{cost(O)} - \text{cost}(S)\)
as follows
Consider client $j$
how much does it contribute to
\[ \sum_{(l, i)} (\text{cost}(S - l + i) - \text{cost}(S)) \]
we consider $i$ when $i = \omega(j)$ and
when $(l=\sigma(j), i)$ is a swap pair (at most two pairs)
(in this second we assume $i \neq w(j)$ )
for other values of $i$ we can ignore $j$'s contribution
since it only helps the inequality (why?)
Analysis

when \( i = \omega(j) \)
\( j \) contributes
\( c_{\omega(j),j} - c_{\sigma(j),j} \)

when \( (l=\sigma(j), i) \) is a swap-pair \( j \) contributes ??

note that \( i \neq \omega(j) \) implies \( \omega(j) \) not captured by choice of
swap pairs (why? think carefully here)

hence we can use \( \pi(j) \) here

\( j \) is assigned to \( \sigma(\pi(j)) \) since \( l=\sigma(j) \) is swapped out. here
we use the fact that \( \sigma(\pi(j)) \) is different from \( \sigma(j) \)
Analysis

$j$ is assigned to $\sigma(\pi(j))$ since $l=\sigma(j)$ is swapped out. Here we use the fact that $\sigma(\pi(j))$ is different from $\sigma(j)$.

Let $r = \sigma(\pi(j))$.

So $j$ contributes $c_{r,j} - c_{\sigma(j)j}$.

Observe that this contribution is non-negative since $\sigma(j)$ is closer to $j$ than $r$ by defn.
Analysis

so j contributes $c_{r,j} - c_{\sigma(j),j}$

observe that this contribution is positive since $\sigma(j)$ is closer to j than r by defn

by triangle ineq

$c_{r,j} \leq c_{\omega(j),j} + c_{\omega(j),\pi(j)} + c_{\pi(j),r}$

note that there can be two (l, i) pairs so total contribution of j is at most

$c_{\omega(j),j} - c_{\sigma(j),j} + 2(c_{\omega(j),j} + c_{\omega(j),\pi(j)} + c_{\pi(j),r} - c_{\sigma(j),j})$
\[ \omega(j) = \omega(\pi(j)) \]

\[ \sigma(\pi(j)) = r \]
Analysis

note that there can be two \((l, i)\) pairs so total contribution of \(j\) is at most

\[
c_{\omega(j),j} - c_{\sigma(j),j} + 2(c_{\omega(j),j} + c_{\omega(j),\pi(j)} + c_{\pi(j),r} - c_{\sigma(j),j})
\]

When we sum up over all \(j\) we notice that the last two terms disapper from the fact the \(\pi\) is a permutation. And also from the fact that \(\omega(\pi(j)) = \omega(j)\) we get the total sum as

\[
cost(O) - cost(S) + 4 cost(O)
\]
**Analysis**

Therefore

\[
\text{cost}(O) - \text{cost}(S) + 4 \text{cost}(O) \geq 0
\]

which implies that

\[
\text{cost}(S) \leq 5 \text{cost}(O) = 5 \text{OPT}
\]