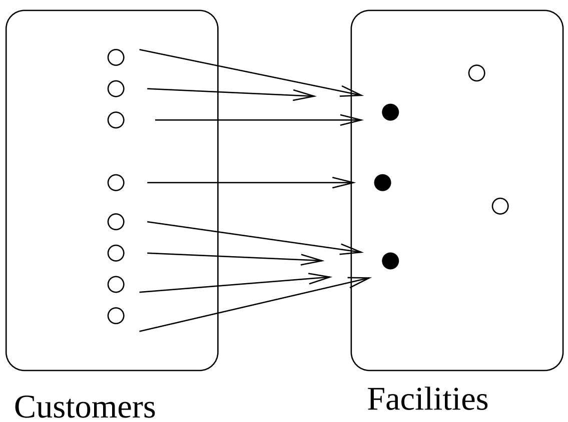
Oblivious medians via online bidding

M. Chrobak, Claire Kenyon, J. Noga and N. Young

k-median

- Choose $\leq k$ facilities: F_k
- $\mathrm{cost}(F_k) = \sum_{c \text{ customer}} d(c, \text{closest facility})$
- Competitive ratio $\mathrm{cost}(F_k)/\mathrm{OPT}_k$.
- Best algorithm: $3 + \epsilon$ [Arya Garg Khandekar Meyerson Munagala Pandit 01 04]

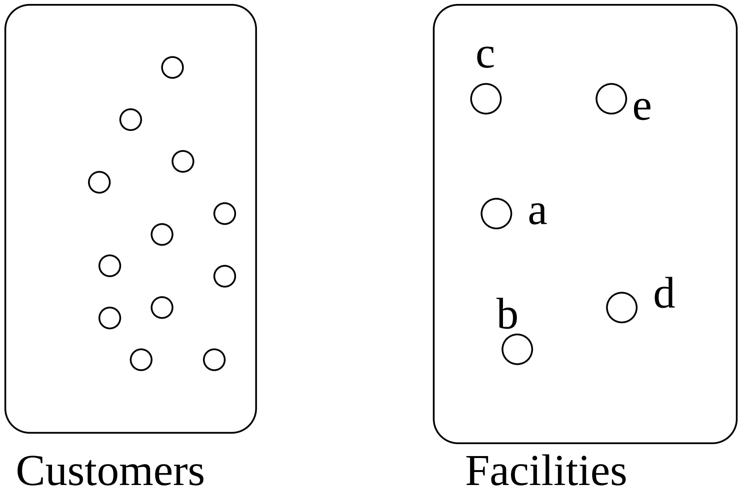


Oblivious median [Mettu Plaxton 99]

- **Algorithm:** Choose $F_1 \subseteq F_2 \subseteq \ldots \subseteq F_n$
- **Adversary:** Chooses k
- **Goal:** Minimize

$$\max_k \text{cost}(F_k)/\text{OPT}_k.$$

$$F_1 = \{a\}, F_2 = \{a, b\}, F_3 = \{a, b, c\}, F_4 = \{a, b, c, d\}, F_5 = \{a, b, c, d, e\}$$

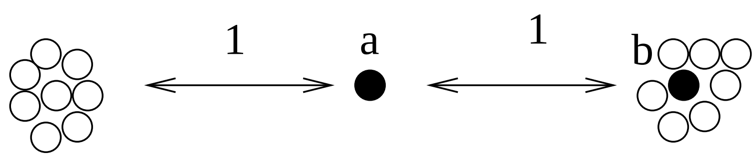


Greedy algorithm

```
F(0) <-- emptyset
For  i <-- 1 to n
    F(i) <-- F(i-1) union {f}, where
     f = greedy choice of next facility to add
```

Greedy fails

- $F_1 = \{a\}$, $f_2 = \{a, b\}$
- $\text{cost}(F_2) = M$
- $\text{OPT}_2 = 1$
- Ratio: M, unbounded



Mettu-Plaxton's algorithm

```
F(0) <-- emptyset
For  i <-- 1 to n
    F(i) <-- F(i-1) union {f}, where
     f = careful choice of next facility to add
```

- Technical algorithm, plus
- Messy analysis, equals
- Competitive ratio around 30

Reverse Greedy [Fiat]

```
F(n) <-- F
For  i <-- n-1 down to 1
    F(i) <-- F(i+1) minus  {f}, where
     f = greedy choice of facility to remove
```

Reverse greedy fails [Chrobak K. Young]

Competitive ratio: $\Omega(\log n / \log\log n)$

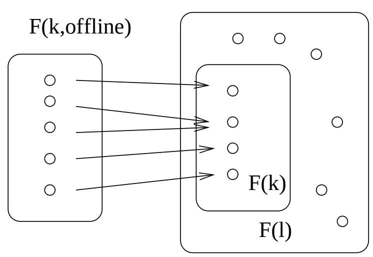
Batch reverse greedy

[Chrobak K. Noga Young, Lin Nagarajan Rajaraman Williamson]

```
Choose a subset K of {1,2,...,n}
F(n) <-- F
For k in K, in decreasing order
     Assume F(l) is already known, l>k, l in K
     F(k) <-- a subset of F(l) of size at most k
     F(k+1),F(k+2),...,F(l-1) <-- F(k)
```

How to choose the subset:

- Let F_k^{off} be a solution to (offline) k-median.
- For $x \in F_k^{\mathrm{off}}$, let $f(x)$ be the point of F_ℓ closest to x.
- Then $F_k = \cup_x \{f(x)\}$



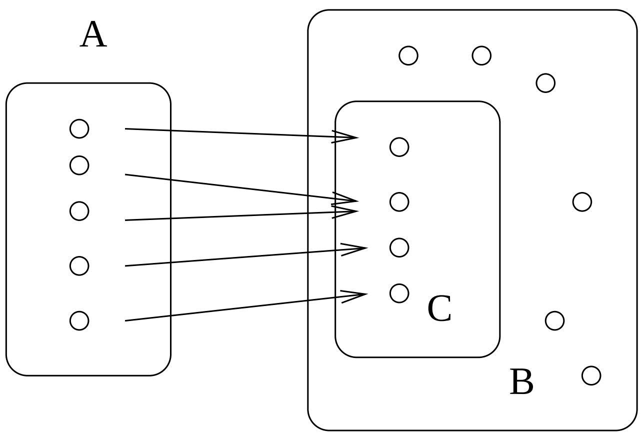
Results

- **($F_k^{\textbf{off}} = \textbf{OPT}_k$)** Competitive ratio of 8 (deterministic) or $2e$ (randomized)
- **($F_k^{\textbf{off}} = (3+\epsilon)$-approximation) Competitive ratio of $24+8\epsilon$ (deterministic) or $6e+2e\epsilon$ (randomized)**
- **Further improvements [LNRW]:** Competitive ratio of $16+O(\epsilon)$ (deterministic) or $4e+O(\epsilon)$ (randomized)

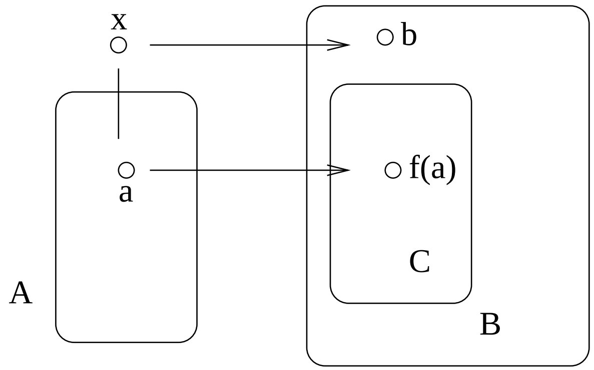
Analyzing one step

Lemma [implicit in Jain-Vazirani]: Let A, B be two sets. Let C be a subset of elements of B containing, for each $a \in A$, an element of B closest to a. Then, for every x,

$$\mathrm{dist}(x, C) \leq 2\mathrm{dist}(x, A) + \mathrm{dist}(x, B).$$



Proof



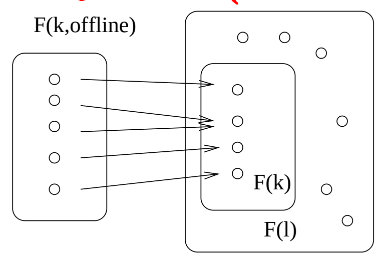
- **a:** point in A closest to x.
- **f(a):** point in B closest to a.
- **b:** point in B closest to x.

$$\text{dist}(x, C) \le \text{dist}(x, f(a)) \le \text{dist}(x, a) + \text{dist}(a, f(a))$$

$$\text{dist}(a, f(a)) \le \text{dist}(a, b) \le \text{dist}(a, x) + \text{dist}(x, b)$$

$$\text{dist}(\mathbf{x}, \mathbf{C}) \le \mathbf{2}\text{dist}(\mathbf{x}, \mathbf{A}) + \text{dist}(\mathbf{x}, \mathbf{B})$$

Analysis (cont'd)



Summing over x:

$$\text{cost}(F_k) \leq 2\text{cost}(F_k^{\text{off}}) + \text{cost}(F_\ell)$$

Iterating:

$$\text{For } \mathbf{k} \in \mathbf{K}: \quad \text{cost}(\mathbf{F_k}) \leq \mathbf{2} \sum_{\ell \geq \mathbf{k}, \ell \in \mathbf{K}} \text{cost}(\mathbf{F}_\ell^{\text{off}})$$

How to choose K

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\text{cost}(F_k^{\text{off}})$	50	45	34	**31**	28	27	20	17	**15**	11	9	**7**	5	**3**	**2**	**1**

$$k = 6 \qquad \text{OPT}_6 = 27 \qquad 16 < 27 \leq 32$$

$$\begin{aligned}
\text{cost}(F_k) &\leq 2(31 + 15 + 7 + 3 + 2 + 1) \\
&\leq 2(32 + 16 + 8 + 4 + 2 + 1) \\
&\leq 2(2F_k^{\text{off}} + F^{\text{off}}k + F_k^{\text{off}}/2 + \cdots) \\
&\leq 8F_k^{\text{off}}
\end{aligned}$$

$\leq (24 + 8\epsilon)$**OPT**$_k$ **if** $F_k^{\textbf{off}}$ **is a** $(3+\epsilon)$**-approximation**

How to choose K (randomized)

Choose x uniformly at random in $[0,1]$.

k	$\cdots$	$\times$	$\cdots$	$\times$	$\cdots$	$\times$	$\cdots$	$\times$
$\mathrm{cost}(F_k^{\mathrm{off}})$	$\cdots$	e^{i+x}	$\cdots\ \mathrm{cost}(F_k^{\mathrm{off}})\ \cdots$	e^{2+x}	$\cdots$	e^{1+x}	$\cdots$	e^{x}

Analysis

$$\mathrm{cost}(F_k) \le 2e^{i+x}\left(1+\frac{1}{e}+\frac{1}{e^2}+\cdots\right) = \frac{2e^{i+x}}{1-1/e}$$

$$\mathbf{E}\left(\frac{\mathbf{e^{i+x}}}{\mathbf{F_k^{\mathrm{off}}}}\right) = \int_0^1 \mathbf{e^z dz} = \mathbf{e-1}$$

$$E(\mathrm{cost}(F_k)) \le \frac{2(e-1)}{1-1/e}F_k^{\mathrm{off}} = 2eF_k^{\mathrm{off}}.$$

Extensions

- λ-relaxed metrics such as squared Euclidian distances
- Oblivious size-competitiveness: $\text{cost}(F_k) \leq \text{OPT}_k$, $|F_k| \leq sk$
- Hierarchical clustering [LNRW]
- Oblivious k-MST, k-vertex cover,... [LNRW]

Relation to previous work

How to choose K: algorithms **look familiar**

Online bidding

- **Universe:** U
- **Adversary:** $T \in U$
- **Algorithm:** sequence of bids $b \in U$, stops as soon as some bid is $\geq T$
- **Cost:** sum of algorithm's bids

Online bidding theorem

The optimal competitive ratio for online bidding equals 4 in the deterministic setting and e in the randomized setting.

Randomized lower bound is new

Recognizing online bidding

- **[Charikar Chekuri Feder Motwani 97], [Dasgupta Long 01]**
  Hierarchical clustering to minimize maximum cluster radius/diameter.
  End of analysis: reduction to online bidding.
- **[Goemans Kleinberg 96]** Minimum latency problem
- **[Chakrabarti Phillips Schultz Shmoys Stein Wein 96]**
  Scheduling algorithms for minsum criteria
- **[Motwani Phillips Torng 94]** Non clairvoyant scheduling
- **[Kao Reif Tate 93]** cow path problem

Randomized lower bound

No randomized algorithm for online bidding can be better than e-competitive.

$U = \{1, 2, \ldots, n\}$

Adversary: T

Algorithm successively bids $b_1 \leq b_2 \leq b_3 \leq \cdots$

Partition of U:

$$[1, b_1] \cup [b_1 + 1, b_2] \cup [b_2 + 1, b_3] \cup \cdots$$

Algorithm pays for bid b iff there is a $t \leq T$ such that $[t, b]$ is in the partition.

LP relaxation

$$\min \beta$$

$$\begin{cases} \sum_{t \le T} b x(t,b) & \le & \beta T & \forall T \\ \sum_{t \le T, b \ge T} x(t,b) & \ge & 1 & \forall T \\ x(t,b) & \ge & 0 & \forall t, b \end{cases}$$

Write dual

Find feasible dual solution with value e

Title