Dynamic Facility Location : Minimizing Sum of Radii

N.K.Blanchard, N.Schabanel

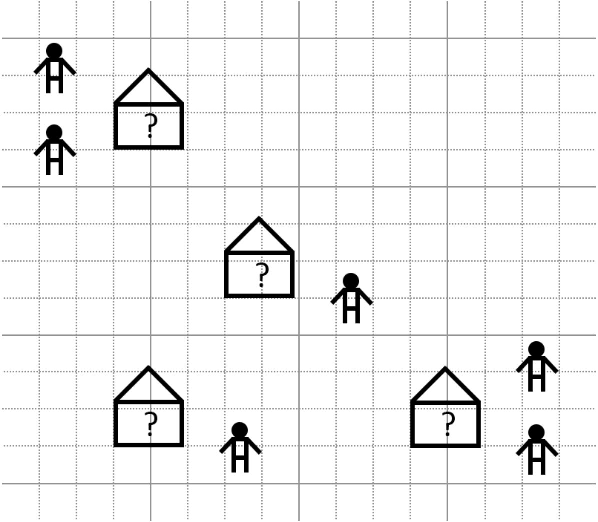
Talk at GT COA
November 23rd, 2015

23 novembre 2015

Plan of the talk

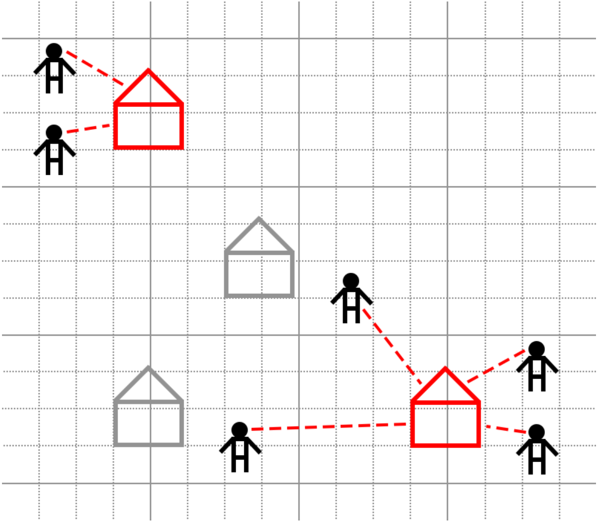
Presentation of the problem

Warehouse Location



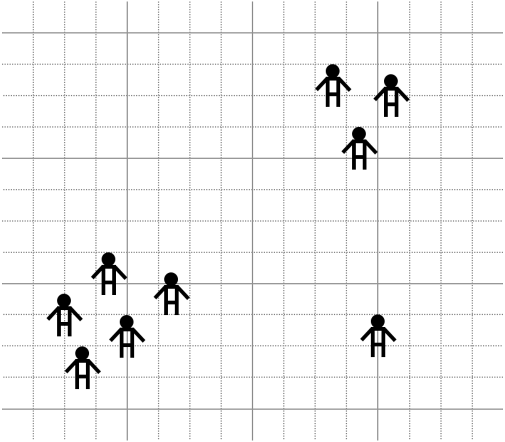
Introduced by Hamburger and Kuehn in 1963

Warehouse Location

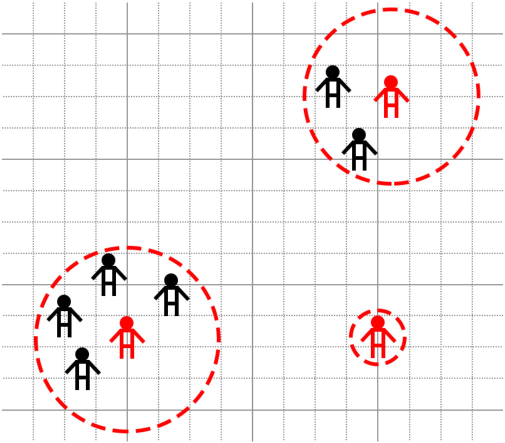


Minimizing opening and connection costs

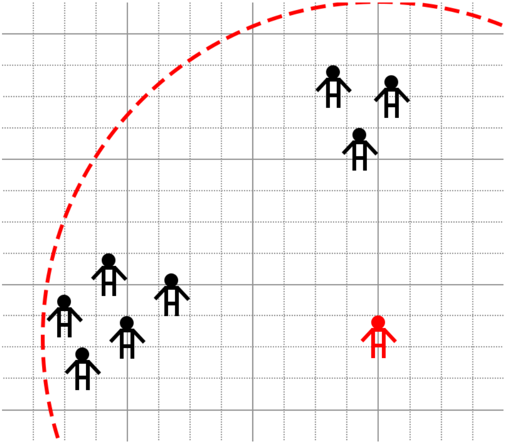
Use for Clustering



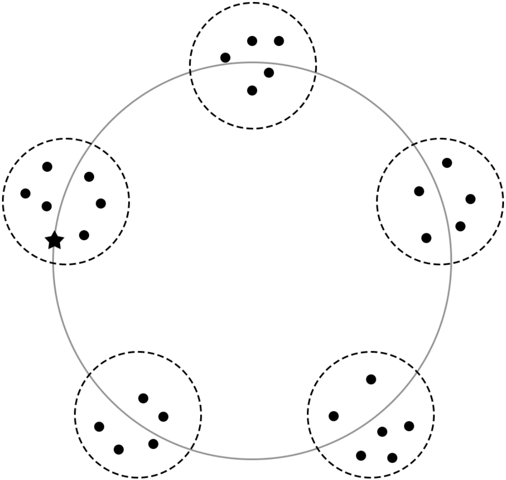
Use for Clustering



Use for Clustering

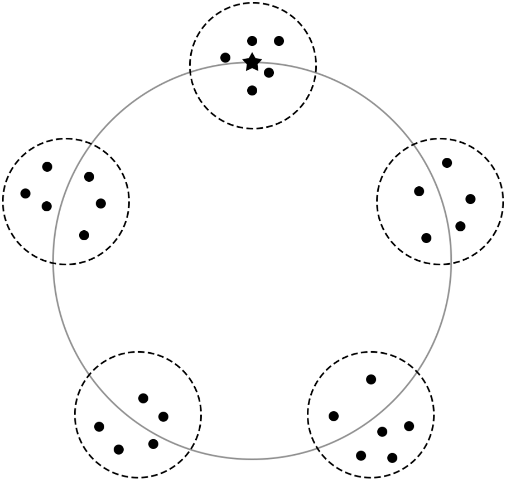


Dynamic aspect



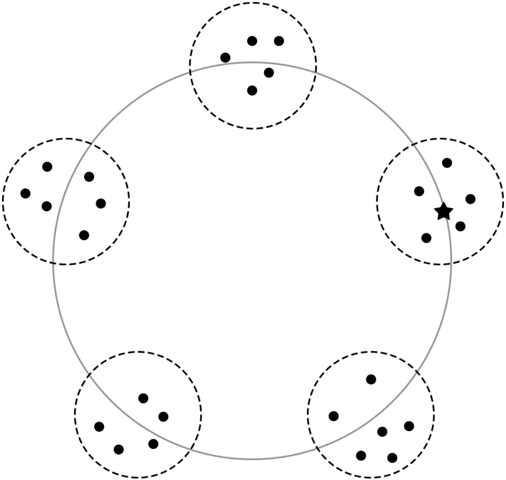
Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Dynamic aspect



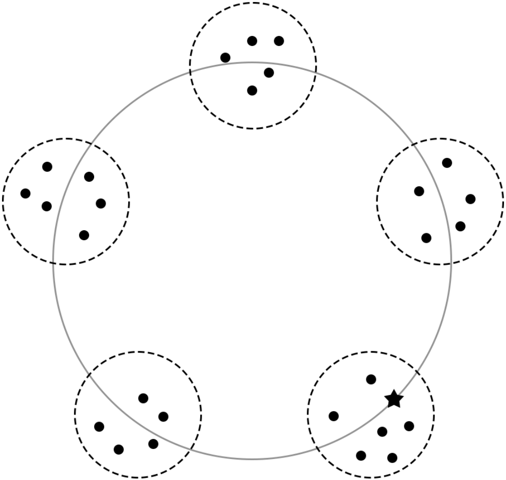
Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Dynamic aspect



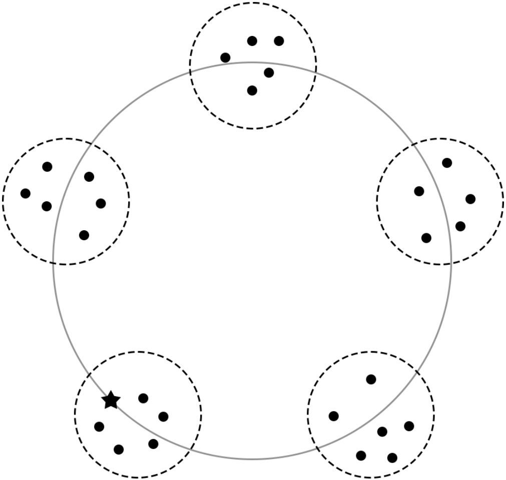
Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Dynamic aspect



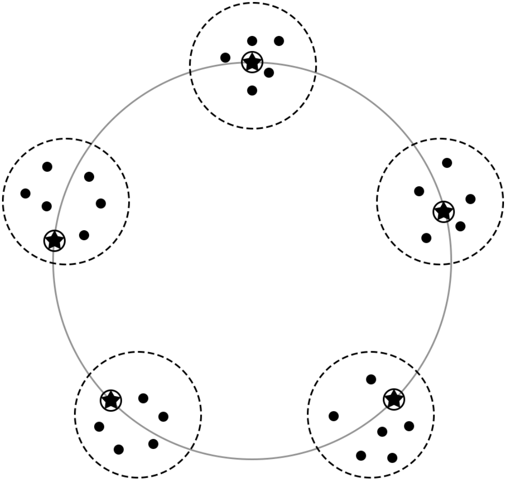
Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Dynamic aspect



Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Dynamic aspect



Dynamic case studied in [EMS 2014] (Introduced by Berman and Odoni in 1982 for stochastic clients)

Definitions

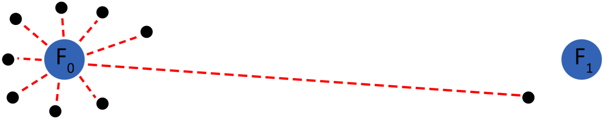
On a en entrée :

- Set $\mathcal{C}$ of n clients
- Set $\mathcal{F}$ of m facilities
- Set $\mathcal{T}$ of t time steps
- Opening cost f and switching cost g
- Distances $d^t(i,j)$ for each (facility, client) pair and each time step

We seek to connect all clients and minimize costs

Problem with the cost function : Sum of distances

Problem with the cost function : Sum of distances



New cost function : Sum of radii

Model studied by Charikar and Panigrahy in 2001 (static case)

New cost function : Sum of radii

Model studied by Charikar and Panigrahy in 2001 (static case)

Reduction from SET-COVER

Reduction from SET-COVER

Théorème

[Dinur - Steurer, 2013] SET-COVER has no $(1-\varepsilon)\log n$ approximation unless $\mathsf{P} = \mathsf{NP}$.

Which gives us :

Théorème

Dynamic Facility Location Minimizing Sum of Radii has no $(1-\varepsilon)\log n$ approximation unless $\mathsf{P} = \mathsf{NP}$.

However this result doesn't hold in the metric case.

Algorithms in the general (non-metric) case

Formulation en Programme Linéaire

Three sets of boolean variables with :

- y_{ir}^t equals 1 iff i is open with radius r at time t
- x_{ij}^t equals 1 iff j is connected to i at time t
- z_{ij}^t equals 1 iff j connected to i between $t-1$ and t

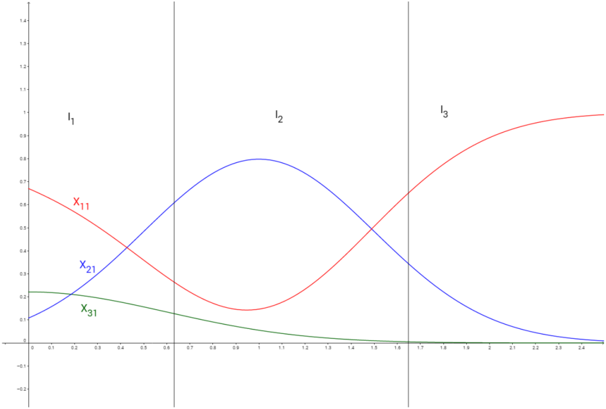
Minimize

$$\sum_{i,t,r} y_{ir}^t \cdot (f + r) \quad + \quad g \sum_{i,j,t} z_{ij}^t$$

With the constraints :

- $\forall j, t \ \ \sum_i x_{ij}^t \geq 1$
- $\forall i, j, t \ \ x_{ij}^t \leq \sum_{r \geq d^t(i,j)} y_{ir}^t$
- $\forall i, j, t \geq 1 \ \ z_{ij}^t \geq x_{ij}^t - x_{ij}^{t-1}$

Preprocessing : choosing when clients switch facilities



Lemme

(EMS) Preprocessing multiplies the cost by at most 2

Opening facilities

Algorithm 1 :

- Repeat $\log(Z) + 1$ times
  - For each facility i, draw a_i uniformly in $[0, 1]$
  - For each time step t, open facility i with the biggest radius R such that

$$a_i \leq \sum_{r \geq R} y_{ir}^t$$

It's equivalent to opening i with a radius distributed along the y_{ir}^t, while maintaining timewise coherence

- Open each facility with the biggest radius among the $\log(Z) + 1$ partial solutions

Covering each client

Interval I of client j isn't covered by facility i if there is $t \in I$ such that

$$a_i \geq \sum_{r \geq d^t(i,j)} y_{ir}^t$$

By the LP's second constraint, for $t \in I$

$$\sum_{r \geq d^t(i,j)} y_{ir}^t \geq \min_{t \in I} x_{ij}^t \stackrel{\text{def}}{=} x_{ij}^I$$

The probability of being covered is then at least

$$1 - \left(\prod_i \left(1 - x_{ij}^I\right) \right) \underset{\text{LP}}{\geq} \frac{1}{2}$$

Repeating $\log(Z) + 1$ times, each client-interval is covered with probability at least $\frac{1}{2}$

Total Cost

- Switching costs at most $Z \cdot g$, by preprocessing at most $2 \cdot OPT$
- Cost expectation by facility equal to $\sum y_{ir}^{t}(r + f)$
- Expecation of cost for partial solutions equal to OPT
- Recombining costs less than summing

Théorème

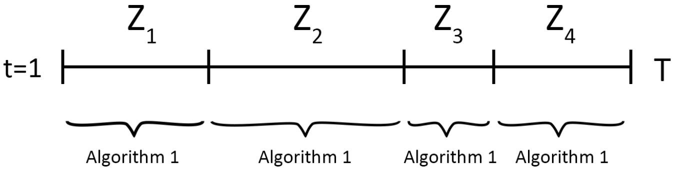
Algorithm 1 is an $O(\log(Z))$ *approximation*

Corollaire

Algorithm 1 is an $O(\log(nT))$ *approximation*

log n-Approximation

- If $Z \leq n^2$ use Algorithme 1
- Else cut time in sequences such that $n \leq Z_i < 2n$ and use 'Algorithm 1 on each sequence



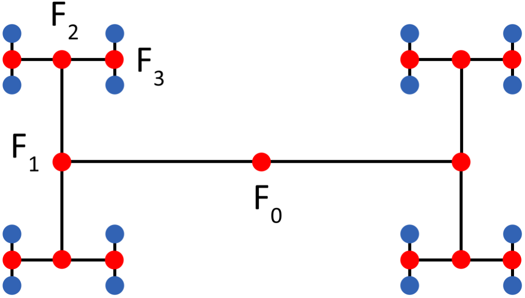
Théorème

Algorithm 2 is an $O(\log n)$ approximation

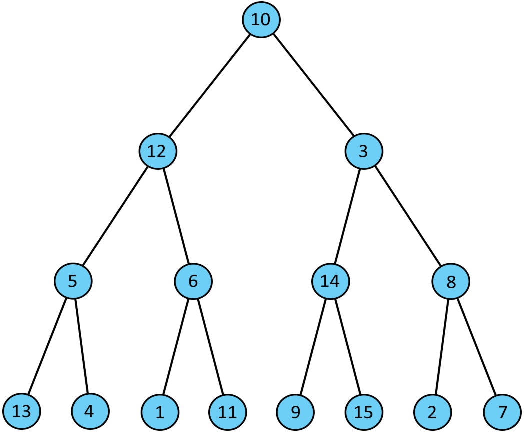
Metric case

ANS Algorithm

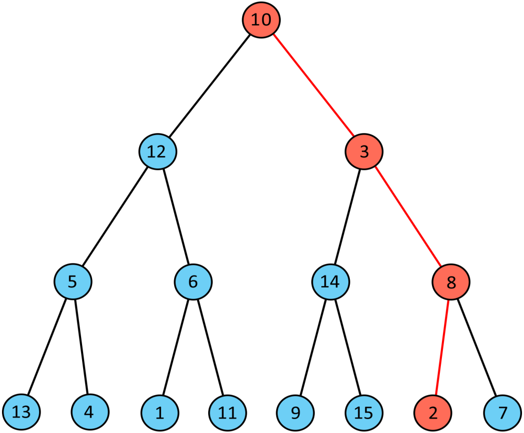
- There is an $O(1)$ approximation for the sum of distances [An, Norouzi-Fard, Svensson, 2014]
- We propose a natural adaptation of this algorithm
- It can be at best a $\Omega(\log \log n)$ approximation
- Counter-example :



Combinatorial Lemma



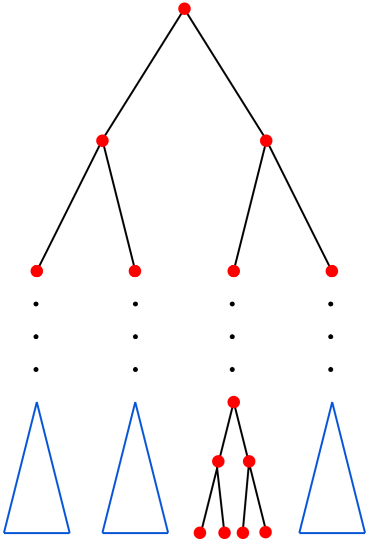
Combinatorial Lemma



Lemme

Finding such a branch has probability at least $\frac{1}{3}$.

Combinatorial Lemma(2)



Conclusion

Conclusion

Our results on Dynamic Facility Location :

	Non-Metric (Algorithm/Hardness)	
	$\sum$ Distances	$\sum$ Radii
Static	$O(\log n)$ / $\Omega(\log n)$	$2 \log n$ / $(1-\varepsilon)\log n$
Dynamic	$O(\log n \cancel{T})$ / $\Omega(\log n)$	$4 \log n$ / $(1-\varepsilon)\log n$

Moreover, modified ANS cannot give better than $\Omega(\log \log n)$ for the sum of radii.

Futur work

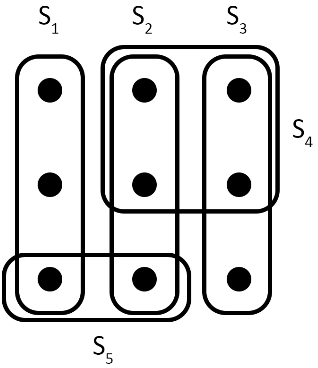
Remaining problems :

- Can we find an upper bound for modified ANS ?
- Can we find a constant approximation algorithm in the metric case ?

Any questions ?

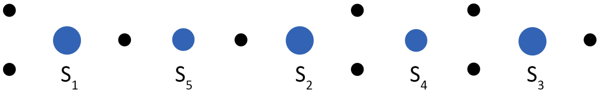
SET-COVER problem

- Set $\mathcal{A}$ with n elements
- Collection $\mathcal{B}$ of subsets of $\mathcal{A}$
- How to cover $\mathcal{A}$ with a minimal cardinality subset of $\mathcal{B}$?



SET-COVER reduction

- A client j for each element $a_j \in \mathcal{A}$
- A facility i for each set $b_j \in \mathcal{B}$
- Set $d(i,j) = 1$ if $a_j \in b_i$ and $d(i,j) = \infty$ otherwise



Preprocessing

- Set $t_0 = 1$
- Greedily find the greatest t_{k+1} such that

$$\sum_i \min_{t \in [t_k, t_{k+1}[} x_{ij}^t \geq \frac{1}{2}$$

- Set for all $[t_k, t_{k+1}]$

$$\widehat{x_{ij}^t} \stackrel{\text{def}}{=} 2 \cdot \min_{t \in [t_k, t_{k+1}[} x_{ij}^t$$

- And

$$\widehat{y_{ir}^t} \stackrel{\text{def}}{=} 2 \cdot y_{ir}^t$$

Proof of the combinatorial lemma

- The root's key is in the top $1/3$ with probability $1/3$
- If we only take an ε proportion of nodes :
  - Each node's key has probability at least $\frac{2}{3} - \varepsilon$ to be smaller than the root's.
  - We can find a path through a subtree of height $\frac{n}{2}$ with constant probability
  - From this node at height $\frac{n}{2}$ we can find an other path of length $\frac{n}{2}$ hence a branch with the property
- We handle dependency and the value of ε by only looking at an arbitrarily small proportion of nodes

Lemme

A good branch exists with probability at least

$$\frac{2 + 27 \cdot \tanh^{-1}\left(\frac{1}{2}\right)}{48} \approx 0.3506$$

Modified ANS Algorithm

- Take an LP solution
- Use the preprocessing
- Draw $a_i \in \mathbb{R}$ for each facility with distribution along $e^{-a_i x}$
- Draw an uniform permutation of clients
- Draw an arrow from client j to facility i with minimal a_i with $x_{ij}^t > 0$
- Draw an arrow from each facility to the client of lowest rank such that $x_{ij}^t > 0$
- Assign each client to the facility in the loop of length2 you get by following the arrows

Experimental results

Series of 10 groups of 10^7 instances for the combinatorial lemma :

Height	5	6	7
Observed probability	0.473	0.453	0.441
Spread over 10 groups	0.0008	0.0018	0.0015

Height	8	9	10
Observed probability	0.431	0.423	0.418
Spread over 10 groups	0.0013	0.0021	0.0017