Dynamic Facility Location : Minimizing Sum of Radii

N.Blanchard

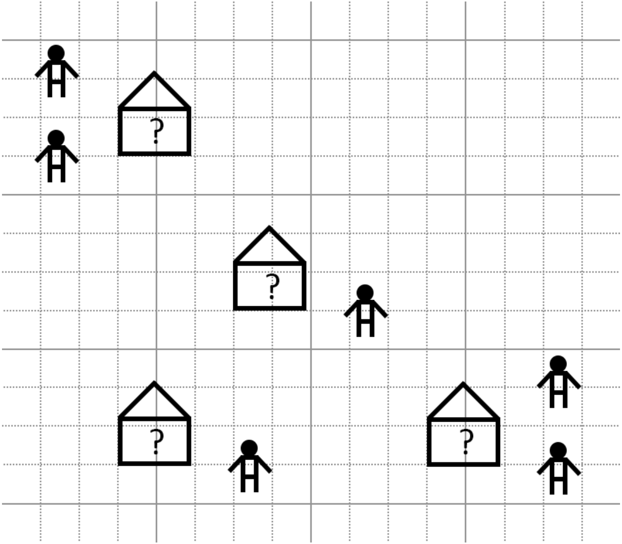
Stage au Laboratoire d'Informatique Algorithmique, Fondements et Applications
Sous la direction de Nicolas Schabanel

3 septembre 2015

Plan de la présentation

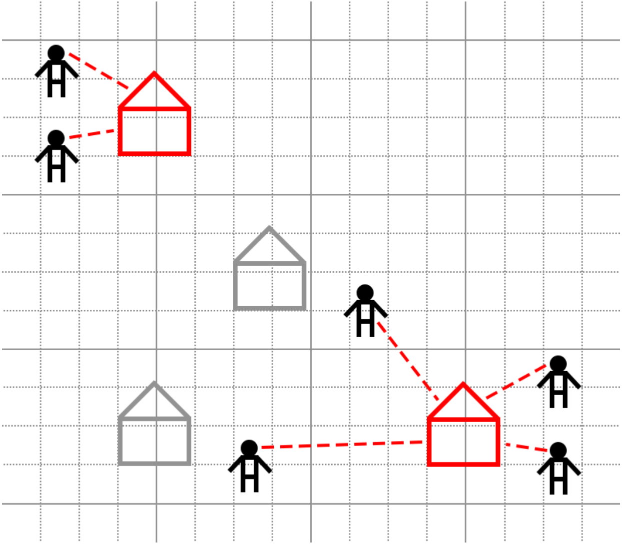
Présentation du problème

Warehouse Location



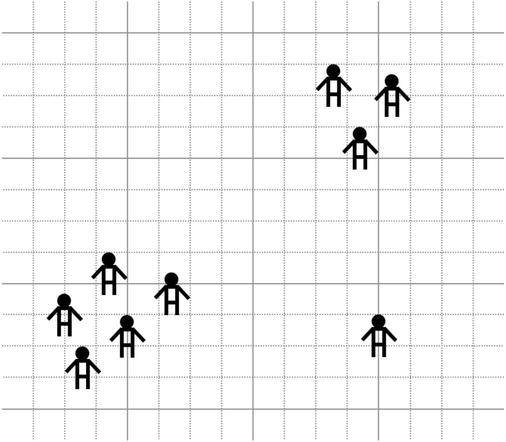
Introduit par Hamburger et Kuehn en 1963

Warehouse Location

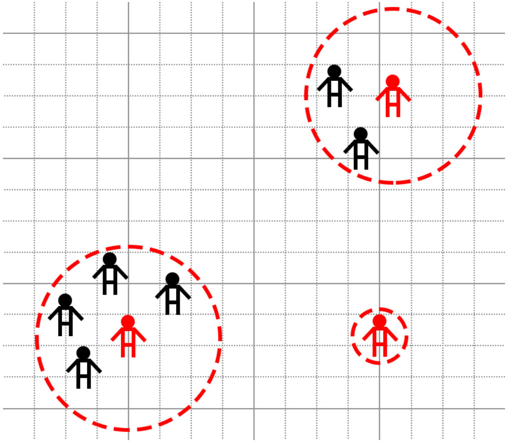


Minimiser les coûts d'ouvertures plus les coûts de connection

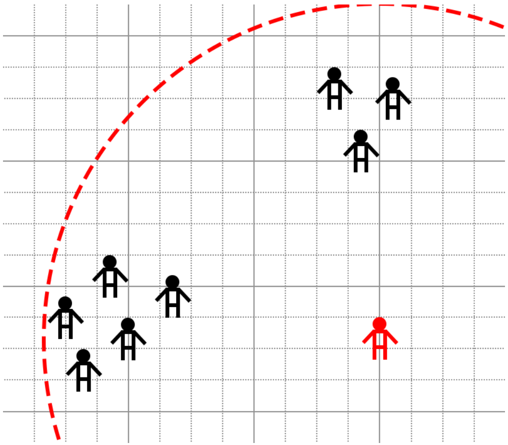
Application au Clustering



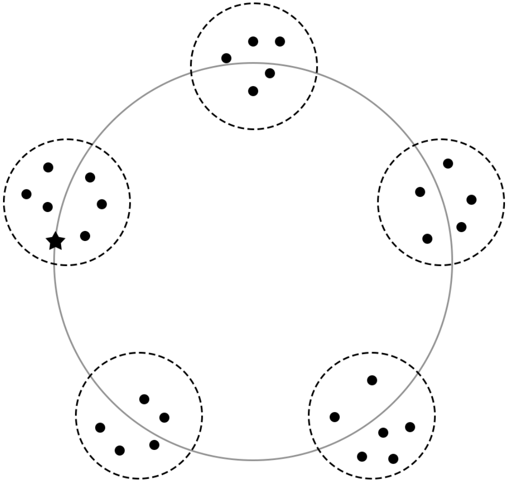
Application au Clustering



Application au Clustering

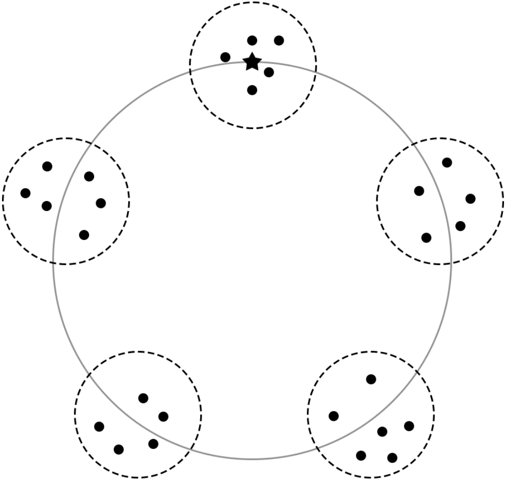


Aspect dynamique



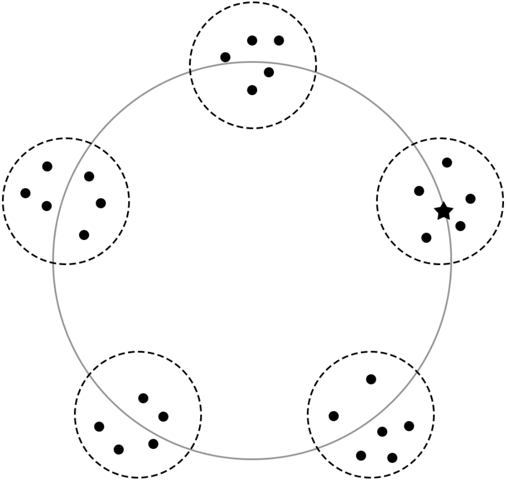
Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Aspect dynamique



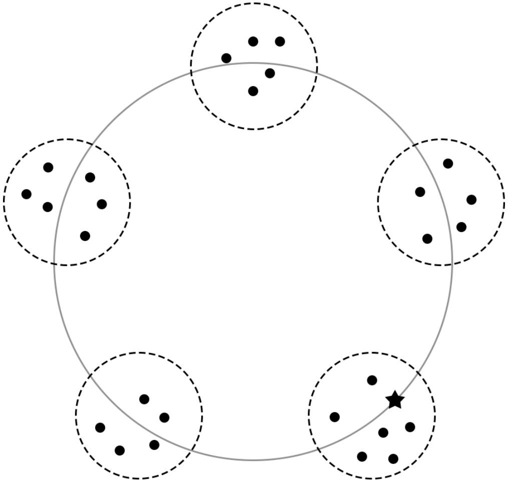
Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Aspect dynamique



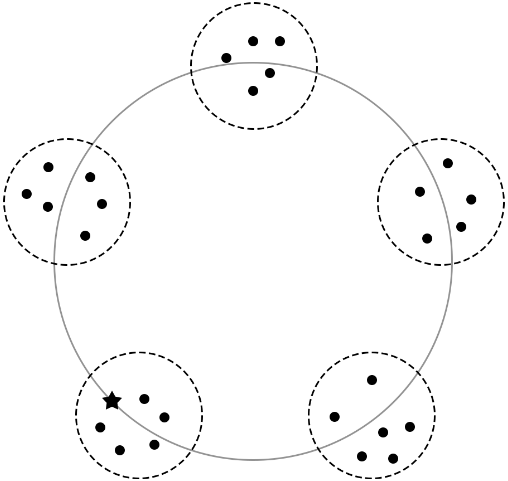
Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Aspect dynamique



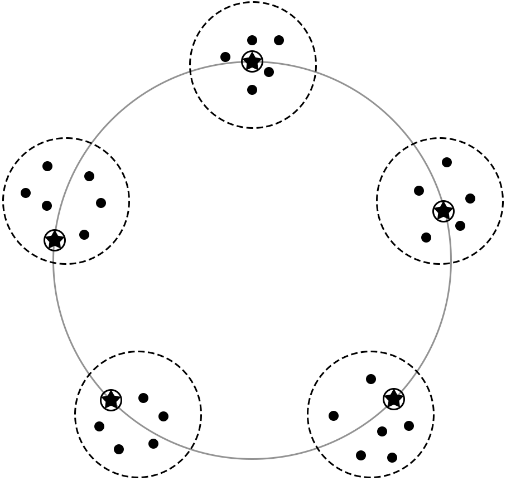
Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Aspect dynamique



Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Aspect dynamique



Cas dynamique étudié dans [EMS 2014] (Introduit par Berman et Odoni en 1982, dans le cadre de clients stochastiques)

Définition du problème

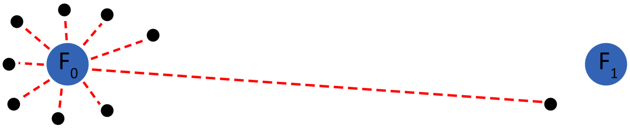
On a en entrée :

- Un ensemble $\mathscr{C}$ de n clients
- Un ensemble $\mathscr{F}$ de m services (facilities)
- Un ensemble $\mathscr{T}$ de t pas de temps
- Un coût d'ouverture f et un coût de changement g
- Des distances $d^t(i,j)$ pour chaque paire (service, client) et chaque pas de temps

On veut connecter tous les clients en minimisant les coûts.

Problème avec l'objectif Somme des distances

Problème avec l'objectif Somme des distances



Nouvelle fonction objectif : Somme des rayons

Modèle étudié par Charikar et Panigrahy en 2001 (en statique)

Nouvelle fonction objectif : Somme des rayons

Modèle étudié par Charikar et Panigrahy en 2001 (en statique)

Réduction de SET-COVER

Réduction de SET-COVER

Théorème

[Dinur - Steurer, 2013] SET-COVER n'admet pas de $(1-\varepsilon)\log n$ *approximation à moins que* $\mathbf{P} = \mathbf{NP}$.

Ce qui permet de montrer :

Théorème

Dynamic Facility Location Minimisant la Somme des Rayons n'admet pas de $(1-\varepsilon)\log n$ *approximation dans le cas non-métrique à moins que* $\mathbf{P} = \mathbf{NP}$.

Dans le cas métrique ce résultat ne tient pas.

Algorithmes dans le cas général (non-métrique)

Formulation en Programme Linéaire

Toutes les variables valent 0 ou 1, avec :

- y_{ir}^t vaut 1 si i est ouverte avec rayon r au temps t
- x_{ij}^t vaut 1 si j est connecté à i au temps t
- z_{ij}^t vaut 1 si j s'est connecté à i entre les temps $t-1$ et t

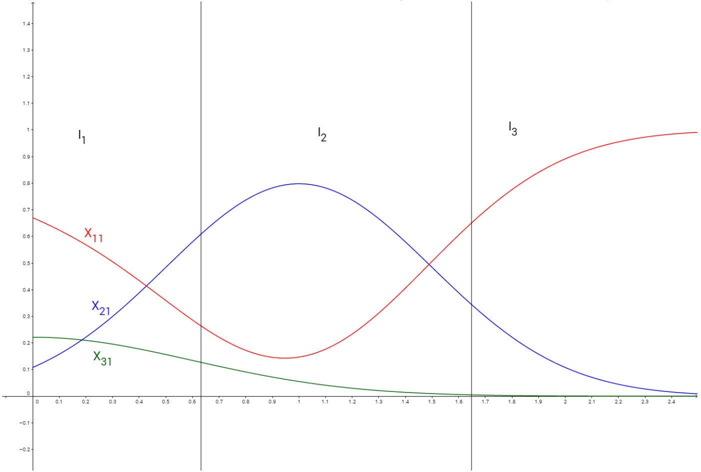
On minimise

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

Avec les contraintes :

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

Preprocessing : décider quand les clients changent de facility



Lemme

(EMS) Le preprocessing double au plus le coût de la solution

Ouverture des facilities

Algorithme 1 :

- On répète ce qui suit $\log(Z) + 1$ fois
  - Pour chaque facility i, on tire une variable a_i uniformément dans $[0, 1]$
  - Pour chaque pas de temps t, on ouvre chaque facility i avec le plus grand rayon R tel que

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

Cela revient à ouvrir i avec un rayon suivant la distribution des y_{ir}^t, en maintenant une cohérence dans le temps

- On ouvre chaque facility avec le plus grand rayon parmi les $\log(Z) + 1$ solutions partielles

Couverture d'un client

Un intervalle I d'un client j n'est pas couvert par une facility i s'il existe $t \in I$

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

Or, par la deuxième contrainte du LP, pour $t \in I$

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

La probabilité d'être couvert par une solution partielle est au moins

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

En répetant $\log(Z) + 1$ fois tous les clients sont couverts sur tous leurs intervalles avec probabilité $\frac{1}{2}$

Coût total

- Coût de changement au plus $Z \cdot g$, par le préprocessing au plus égal à $2 \cdot OPT$
- Espérance de coût par facility égale à $\sum y_{ir}^{t}(r+f)$
- Espérance des coûts des solutions partielles égale à OPT
- Recombiner coûte moins que la somme des coûts partiels

Théorème

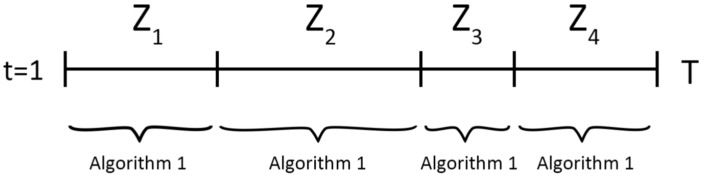
L'algorithme 1 est une $O(\log(Z))$ approximation

Corollaire

L'algorithme 1 est une $O(\log(nT))$ approximation

$\log n$-Approximation

- Si $Z \leq n^2$ lancer l'Algorithme 1
- Sinon découper en séquences telles que $n \leq Z_i < 2n$ et lancer l'Algorithme 1 sur chaque séquence



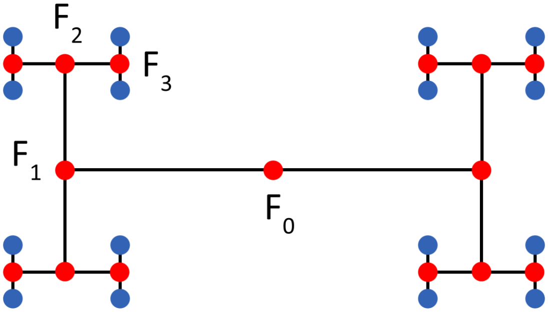
Théorème

L'Algorithme 2 est une $O(\log n)$ *approximation.*

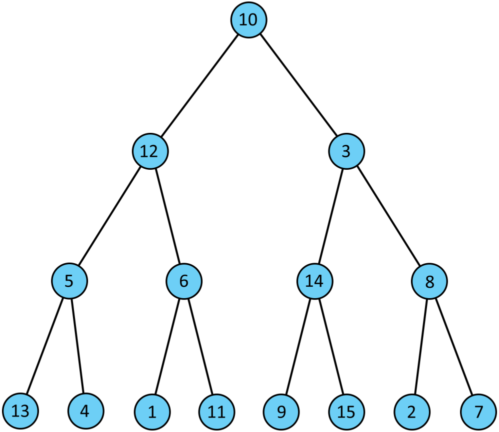
Cas métrique

Algorithme ANS

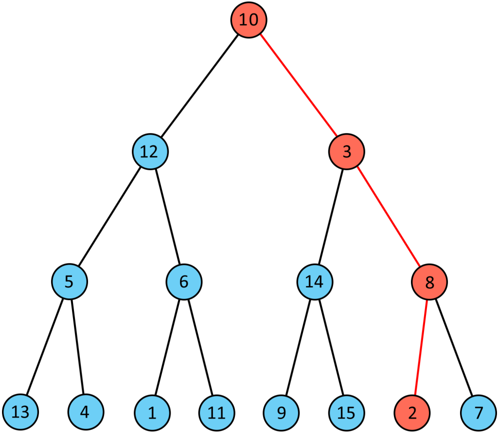
- Il existe une $O(1)$ approximation pour la somme des distances [An, Norouzi-Fard, Svensson, 2014]
- On a proposé une adaptation naturelle de cet algorithme
- C'est hélas au mieux une $\Omega(\log\log n)$ approximation
- Contre-exemple :



Lemme combinatoire



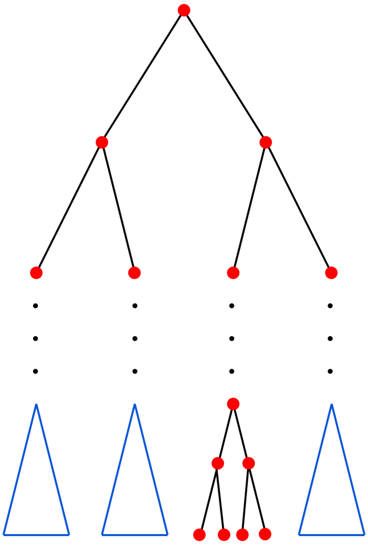
Lemme combinatoire



Lemme

La probabilité de trouver une telle branche est au moins $\frac{1}{3}$.

Lemme combinatoire (2)



Conclusion

Conclusion

Nos résultats sur le Dynamic Facility Location :

	Non-Métrique (algorithme/difficulté)	
	$\sum$ Distances	$\sum$ Rayons
Statique	$O(\log n)$ / $\Omega(\log n)$	$2\log n$ / $(1-\varepsilon)\log n$
Dynamique	$O(\log n\cancel{T})$ / $\Omega(\log n)$	$4\log n$ / $(1-\varepsilon)\log n$

De plus, notre ANS modifié ne peut pas donner une approximation meilleure que $\Omega(\log\log n)$ pour la somme des rayons.

Travail futur

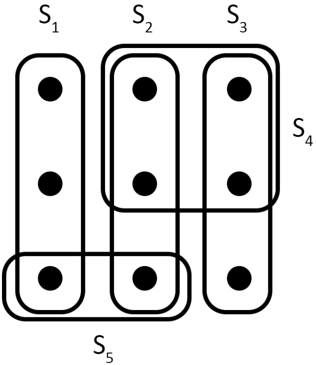
Questions restantes :

- Peut-on prouver une borne supérieure pour le facteur d'approximation d'ANS modifié ?
- Peut-on trouver un algorithme en approximation constante pour le cas métrique ?

Y-a-t'il des questions ?

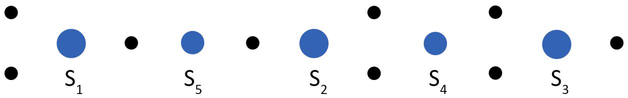
Problème SET-COVER

- Un ensemble $\mathscr{A}$ à n éléments
- Une collection $\mathscr{B}$ de sous-ensembles de $\mathscr{A}$
- Comment couvrir $\mathscr{A}$ avec un sous-ensemble de $\mathscr{B}$ de cardinalité minimale ?



Réduction de SET-COVER

- On prend un client j pour chaque élément $a_j \in \mathscr{A}$
- On prend une facility i pour chaque élément $b_j \in \mathscr{B}$
- On fixe $d(i,j) = 1$ si $a_j \in b_i$ et $d(i,j) = \infty$ sinon



Preprocessing

- On fixe $t_0 = 1$
- On cherche de manière gloutonne le plus grand t_{k+1} possible tel que

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

- Enfin on fixe pour tout intervalle $[t_k, t_{k+1}]$

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

- Et

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

Preuve du lemme combinatoire

- La racine a une clé dans le $1/3$ supérieur avec probabilité $1/3$
- Si on regarde une proportion ε des noeuds :
  - Chaque noeud a donc probabilité $\frac{2}{3} - \varepsilon$ d'être inférieur à la racine
  - On peut trouver un chemin à travers un arbre de hauteur $\frac{n}{2}$ avec probabilité constante
  - À partir de ce noeud à hauteur $\frac{n}{2}$ on peut trouver un deuxième chemin de longueur $\frac{n}{2}$ et donc une branche satisfaisant la propriété
- On gère la dépendance et donc la valeur de ε en ne regardant qu'une petite proportion des noeuds.

Lemme

Une bonne branche existe avec probabilité au moins

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

Algorithme ANS modifié

- On prend une solution au LP
- On utilise le preprocessing
- On tire une variable $a_i \in \mathbb{R}$ par facility distribuée selon $e^{-a_i x}$
- On tire une permutation uniforme des clients
- On met une arête de chaque client à la facility de a_i minimal tel que $x_{ij}^t > 0$
- On met une arête de chaque facility au client d'ordre minimal tel que $x_{ij}^t > 0$
- On assigne chaque client à la facility dans la boucle de longueur 2 qu'on obtient en suivant les arêtes

Résultats expérimentaux

Résultats d'une série de 10 groupes de 10^7 instances pour le lemme combinatoire :

Hauteur	5	6	7
Probabilité empirique	0.473	0.453	0.441
Écart parmi 10 tests	0.0008	0.0018	0.0015

Hauteur	8	9	10
Probabilité empirique	0.431	0.423	0.418
Écart parmi 10 tests	0.0013	0.0021	0.0017