

Algorithme Li et al. : ensemble dominant connexe glouton

Alexandre Fernandez & Sylvain Ung

22 octobre 2017



Rapport de Recherche

Table des matières

1	Introduction	4
2	Algorithme	5
2.1	Construction de l'ensemble stable maximal	5
2.2	Construction de l'ensemble dominant connexe	7
3	Résultats	11
3.1	Tests	11
3.2	Discussion	13
4	Conclusion	14

Algorithme Li et al. : ensemble dominant connexe glouton

Résumé : Il va s'agir dans ce rapport de recherche d'étudier une implémentation efficace d'un algorithme répondant au problème de l'ensemble dominant connexe qui est très largement utilisé pour modéliser un réseau sans-fil. L'algorithme étudié est celui de Li et al [3] qui propose une approche gloutonne de la solution. L'objectif étant de minimiser cet ensemble afin de réduire, en pratique, les coûts de maintenance par exemple.

Mots-clés : Théorie des graphes, ensemble dominant, ensemble stable maximum, connexe, NP-complet, algorithme glouton, Li et al.

1 Introduction

Étant donné un graphe $G = (V, E)$ avec V (respectivement E) l'ensemble des sommets (respectivement l'ensemble des arêtes), l'ensemble dominant connexe¹ de G est le sous-ensemble $D \subseteq V$ tel que $\forall v \in V, v \in D \vee v$ est un voisin d'un élément de D et tel que $G[D]$, le sous-graphe induit par D est connexe.

Cette définition nous indique que le CDS n'est pas unique pour un graphe donné (sauf cas dégénérés) et qu'il en existe plusieurs de tailles différentes. Cela nous amène au problème que nous allons traiter : la recherche du plus petit CDS. C'est un problème NP-difficile ce qui signifie que nous ne pourrions pas calculer le plus petit, autrement dit la meilleure solution, en un temps raisonnable. C'est pour cela que nous allons présenter dans ce rapport un algorithme donnant une solution qui s'en rapproche.

Ainsi, l'algorithme Li et al. propose une approche gloutonne à la résolution de ce problème dont le principe de résolution se fait étape par étape. En effet, à chaque étape un optimum local est choisi afin d'arriver à un optimum global à la fin. Cet algorithme peut être décomposé en 2 grandes étapes :

1. Construire l'ensemble stable maximal²
2. Connecter les points de l'ensemble précédemment construit entre eux par des points qui n'y appartiennent pas, ceux-ci sont appelés *nœuds de Steiner*

Dans notre démarche, nous présenterons tout d'abord les idées théoriques qui se trouvent derrière l'algorithme puis nous détaillerons l'implémentation que nous en avons fait, éventuellement des choix et des modifications pris par rapport l'algorithme original. Ensuite, nous présenterons les conditions et résultats des tests obtenus suivi d'une discussion sur la pertinence d'un tel algorithme.

Enfin, nous conclurons sur les enjeux et problématiques de l'ensemble dominant connexe dans les graphes géométriques, également appelés graphes de disques.

1. CDS (Connected Dominating Set)

2. MIS (Maximum Independent Set), nous en donnerons la définition dans la section 2.1

2 Algorithme

2.1 Construction de l'ensemble stable maximal

La première étape de l'algorithme consiste en la construction du MIS dont la définition est la suivante et qui possède des propriétés intéressantes pour la suite :

Définition 2.1.1 Soit $G = (V, E)$ un graphe, l'ensemble stable maximal est le plus grand sous-ensemble $S \subseteq V$ tel que le sous-graphe $G[S]$ induit par S ne contient pas d'arêtes.

Lemme 2.1.1 Dans tout graphe géométrique, la taille de ses ensembles stables maximaux est majorée par $3.8\text{opt} + 1.2$ où opt est la taille de l'ensemble connexe dominant minimum.

Lemme 2.1.2 Toute paire de sous-ensembles complémentaires du MIS a exactement une distance de deux sauts.

Lemme 2.1.1 garantit que le ratio de la solution de cet algorithme par rapport à la solution optimale est bien $4.8 + \ln 5$. La validité de ce lemme dépend beaucoup de nos choix d'implémentation pour les algorithmes de construction MIS. Ces algorithmes étant prévus pour fonctionner de façon distribuée, nos implémentations, non distribuées, ne se comporte donc pas exactement comme ceux-ci.

Lemme 2.1.2 est d'une importance majeure pour la validité de la solution. En effet, si une paire de sous-ensembles complémentaires du MIS a une distance inférieure à deux sauts, le MIS sera plus grand que prévu. On risque donc dans ce cas d'obtenir un résultat final plus grand. À l'inverse, si la distance entre deux sous ensembles complémentaires du MIS est de plus de deux sauts. L'algorithme, essayant de relier ces sous-ensembles du MIS en colorant des nœuds contenant au minimum un voisin dans chacun de ces sous ensembles, ne pourra pas relier ces deux sous-ensembles. L'ensemble obtenu sera constitué d'au moins 2 composantes connexes correspondant aux deux ensembles complémentaires du MIS ayant une distance de plus de deux sauts.

Afin de construire un tel MIS qui répond correctement à ces propriété nous nous sommes appuyés une des références [1] cités par les auteurs de l'article. Comme précédemment dit, nous avons dû adapter l'algorithme pour un fonctionnement dans une architecture non distribuée qui faisait passer des messages entre les différents nœuds afin de connaître l'état du réseau. Dans notre cas, nous nous sommes servis d'un système de marquage pour connaître l'avancement de la construction du MIS :

- **noir** : nœud dominant, appartient au MIS
- **gris** : nœud dominé, voisin d'un nœud dominant, n'appartient pas au MIS
- **blanc** : nœud encore non traité par l'algorithme
- **blanc actif** : état particulier d'un nœud potentiellement prêt à devenir dominant

Pour illustrer le fonctionnement de notre algorithme basé sur [1] nous allons prendre l'exemple ci-dessous :

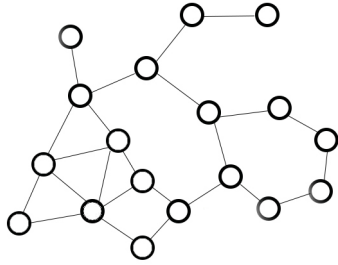


FIGURE 1 – MIS : état initial

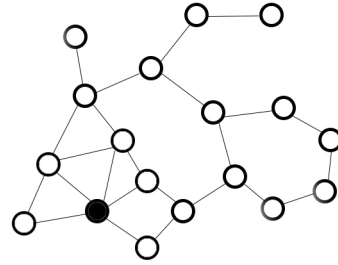


FIGURE 2 – MIS : leader

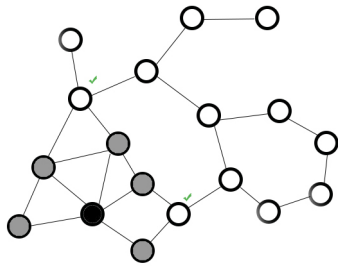


FIGURE 3 – MIS : domination

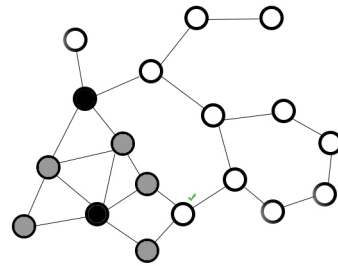


FIGURE 4 – MIS : élection

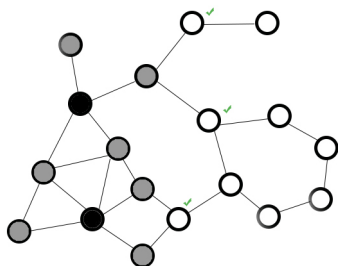


FIGURE 5 – MIS : domination

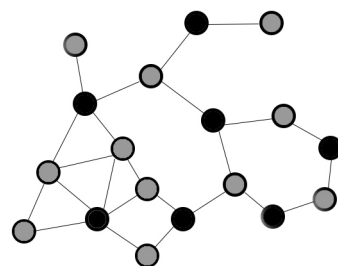


FIGURE 6 – MIS : état final

L'algorithme commence par choisir un tout premier nœud comme leader, en l'occurrence nous avons fait le choix de prendre le sommet de plus haut degré ce qui semble être pour nous un choix justifié puisque une plus large zone du graphe peut être couverte par ce seul point (figure 2). Ce nœud hôte est ainsi coloré en noir pour marqué son appartenance au MIS et tous ses voisins en gris pour indiquer leur domination par ce dernier. (figure 2).

2.2 Construction de l'ensemble dominant connexe

L'algorithme construit le CDS tel qu'il est décrit dans le papier :

```

compute a MIS satisfying lemmas 1 and 2.
color MIS nodes to black.
color other nodes to grey.
for i = 5; 4; 3; 2 do
    while there exists a grey node
        adjacent to at least i black
        nodes in different black-blue
        components
    do change its color from grey to
        blue;
return all blue nodes.

```

La principale difficulté au niveau de l'implémentation de l'algorithme est au sujet des *black-blue components*. Il faut rapidement identifier à quel *black-blue component* appartient un point noir. De plus quand un nouveau point bleu est choisi, il provoque la fusion de tout ses *black-blue components* voisins. La fusion de *black-blue components* (*BBC*) peut être coûteuse en fonction de l'implémentation. Une implémentation naïve consisterait à utiliser un ensemble de listes pour représenter les *BBC*, chaque liste contenant les points appartenant à un *BBC*. Considérons le cas au pire cas de toutes les opérations fusions de *BBC*.

Considérons le cas dégénéré d'un chemin de points de degré 2. Dans ce cas, chaque point gris, a exactement deux voisins noirs. Pour chaque point gris, on va réaliser une unique fusion (ie : fusionner uniquement 2 *BBC*). Si l'on considère, pour simplifier l'analyse que l'ordre de choix des points gris est tel que on va toujours fusionner un même *BBC* avec un *BBC* ne contenant qu'un seul point, dans le cas pire, on vide toujours la liste la plus grande dans la liste ne faisant qu'un en taille. On obtient le nombre d'opération ci dessous :

$$1 + 2 + 3 + 4 + \dots + n = \frac{n(n+1)}{2}$$

avec $n = \text{taille}(\text{MIS})$

On obtiens un coût en $O(n^2)$ pour les fusions.

Pour les recherches, on va faire un parcours de tout les *BBC* pour y trouver les voisins d'un point gris, ce qui, au pire cas où il faut parcourir tout les *BBC* pour y trouver les voisins, sachant que l'ensemble des points dans les *BBC* correspond à la taille du MIS, on obtiens le nombre de comparaisons suivant :

$$n \times m$$

avec $n = \text{taille}(\text{MIS})$ et $m = \text{taille}(\text{ptsGris})$, $\text{taille}(\text{ptsGris}) = \text{taille}(\text{points}) - \text{taille}(\text{MIS})$.

Ces deux opérations sont grossièrement au pire cas quadratiques et il convient de les optimiser. Pour cela, nous allons utiliser une structure de Disjoint-Set pour réaliser des opérations UNION et FIND pour représenter les BBC.

Cette structure organise différents ensembles disjoints sous la forme d'une forêt d'arbres dont les noeuds pointent sur leur parents. Au début tout les BBC sont initialisés à un point noir et sont disjoints. Puis, lorsque l'on veut faire l'union de deux BBC, on cherche leur racines et si elles sont différentes, un parent commun est créé pour les deux racines : opération UNION.

Lorsque l'on veut vérifier si deux points noirs appartiennent au même BBC, il suffit de remonter leurs parents jusqu'à la racine. Si deux points ont la même racine alors ils sont dans le même ensemble. Au pire cas, avec une implémentation naïve des Disjoint Set, les arbres peuvent prendre la forme d'un peigne et le coût au pire cas des opérations est $O(n)$. Ce qui nous ramènerait à des complexités quadratiques pour la recherche et la fusion au cours de tout l'algorithme.

Cependant, deux optimisations simples sont très efficaces avec les UNION-FIND :

- la path compression : lors d'une opération recherche, on remonte les éléments traversés comme fils de la racine.
- l'union par rang : lors d'une fusion, si les deux arbres ont la même taille, on crée une nouvelle racine, commune, pour ces deux arbres, cependant quand un arbre est moins haut que l'autre, on peut seulement ajouter comme parent de la racine de l'arbre le moins haut, la racine de l'arbre le plus haut.

Avec ces deux optimisations le coût amorti des deux opérations est $O(a(n))$ avec a l'inverse de la fonction d'Ackermann, une fonction ayant une croissance si lente qu'on peut la considérer comme une constante inférieure à 5 pour quelque situation.

Il est justifié de parler de coût amorti dans notre cas, car on réalisera toujours des opérations find (réalisant la path compression) pour trouver les black blue components différents voisins d'un point gris, avec de réaliser des opérations union pour fusionner tout les bbc voisins d'un point nouvellement bleu.

MakeSet crée des sets singletons pour l'initialisation :

```
function MakeSet(x)
  if x is not already present:
    add x to the disjoint-set tree
    x.parent := x
    x.rank  := 0
```

Find implémente la recherche de la racine avec la path compression :

```
function Find(x)
  if x.parent != x
    x.parent := Find(x.parent)
```

```
return x.parent
```

Union fait l'union par rang

```
function Union(x, y)
  xRoot := Find(x)
  yRoot := Find(y)

  // x and y are already in the same set
  if xRoot == yRoot
    return

  // x and y are not in same set, so we merge them
  if xRoot.rank < yRoot.rank
    xRoot.parent := yRoot
  else if xRoot.rank > yRoot.rank
    yRoot.parent := xRoot
  else
    //Arbitrarily make one root the new parent
    yRoot.parent := xRoot
    xRoot.rank := xRoot.rank + 1
```

L'utilisation des Disjoint Set a d'autres avantages, l'algorithme est beaucoup plus simple à écrire une fois cette structure implémentée. De plus cette structure permet facilement de vérifier la connexité d'un graphe. Elle nous sera donc utile pour implémenter la fonction de vérification `isValid` et pour générer des graphes géométriques connexes.

Nos structures :

```
class NodeVertexDS {
  Point p;
  Color c;

  ArrayList<NodeVertexDS> neighbors;
  DisjointSetElement<NodeVertexDS> disjointsetelement;

  //d'autres attributs sont utilises pour definir la structure d'arbre
  //necessaire pour l'algorithme du MIS 2
  //d'autres attributs sont utilises pour l'algorithme du MIS 1
}

class DisjointSetElement<T> {
  int index; //unique index
  T data;
  DisjointSetElement<T> parent;
  int rank;

  //toutes les methodes pour implementer les disjoint set : en
  //particulier find et union
}
```

Pseudo-code de notre implémentation :

```

//construire la structure du graphe avec la liste de points en entree et
    edgeTreshold
graph : ArrayList<NodeVertexDS> = makeGraph(points);
MIS : ArrayList<NodeVertexDS> = calculMIS(graph);
grayNodes : ArrayList<NodeVertexDS> = { graph } \ { MIS }

for(n : MIS) { n.color = BLACK; n.disjointsetelement = new
    DisjointSetElement<NodeVertexDS>(n); }

for( i from 5 to 2 ) {
    cont = true;
    while(cont) {
        newblue = null;
        for( gray in grayNodes ) {
            if( gray.degree() < i ) continue;

            //assuming BlkNeighbors returns the black neighbors of a node
            blackneighbors : ArrayList<NodeVertexDS> = BlkNeighbors(gray);
            if( blackneighbors.size() < i ) continue;

            //assuming map is the functional transformation on list
            rootsblackneighbors : ArrayList<DisjointElement> = List.map(
                blackneighbors, (x) => x.disjointsetelement.find() )

            int nbDifferentBBC = new HashSet(rootsblackneighbors).size();
            if( nbDifferentBBC < i ) continue;

            gray.color = BLUE;
            newblue = GRAY;

            for( i from 1 to rootblackneighbors.size() -1)
                rootblackneighbors[i].union(rootblackneighbors[0]);
            break;
        }
        if ( newblue != null ) grayNodes.remove(newblue);
        else cont = false;
    }
}
return BLACK and BLUE nodes;

```

3 Résultats

3.1 Tests

Pour tester notre algorithme nous avons eu besoin de deux choses : La méthode `isValid()`, qui vérifie la validité de la solution calculée par notre algorithme, voici son pseudo-code :

```

isValid(ArrayList<Point> graph, ArrayList<Points> cds, int edgeTreshold)
{
    valid = true;
    //build the graph structure of the cds
    ArrayList<NodeVertexDS> graphcds = graph(cds);

    //compute connex components
    for( v in graphcds ) {
        for( vn in v.neighbors ) {
            v.disjointsetelement.union(vn.disjointsetelement)
        }
    }

    compconnex = new HashSet<DisjointSetElement>();
    for( v in graphcds ) { compconnex.add(v.disjointsetelement.find()); }

    if(compconnex.size() > 1) { print("Error on connexity : " +
        compconnex.size()); valid = false; }

    rest = points.clone();
    rest.removeAll(cds);

    //remove all neighbors of elements of cds from rest
    removeneighbors(rest, cds, edgeTreshold);

    if(rest.size() > 0) { print("Error dominating : " + rest.size());
        valid = false; }

    return valid;
}

```

Nous avons aussi eu besoin de générer des tests. Pour cela il conviens de générer des graphes géométriques de différentes tailles et avec des thresholds différents. Voici le pseudo code du générateur :

```

generateGraph(width, heigth, nb, edgeTreshold) :
    result : ArrayList<Point>;
    while result.size() != nb :
        add random points in result until result.size() == nb
        compute connected components of result
        if there is multiple connected components :
            keep only the biggest connected components in points (remove
            the points that are in smaller components)

```

Nous avons ensuite établi plusieurs bases de test

(S'il te plaît sylvain fait des tableaux :p)

Base 1 : nb points : 100, 500, 1000, 5000, 10000 width and height : 1000, 1000, 1000, 1000 edgeThreshold : 50, 50, 50, 50, 50

Base 2 : nb points : 100, 500, 1000, 5000, 10000 width and height : 100, 500, 1000, 5000, 10000 edgeThreshold : 5, 25, 50, 250, 500

Base 3 : (pour cette base on souhaitait fixer edgeThreshold, cependant, un edgeThreshold de 50 pour un test de 10000 points dans un espace de 10000*10000 est trop long à générer) à la place nous avons choisi un compromis. soit $edeThreshold = (1/36)*nbPoints + 200/9$ nb points : 100, 500, 1000, 5000, 10000 width and height : 100, 500, 1000, 5000, 10000 edgeThreshold : 25, 36, 50, 161, 300

Une base 4 qui génère des points dans un espace beaucoup plus large que haut serait intéressante (à voir demain).

Tout ces tests seront effectués sur les deux version de l'algorithme (avec MIS1 et MIS2)

Résultats faire des graphes de taille du cds en fonction du nb de points et de temps en fonction du nombre de points pour MIS 1 et 2 4 graphes par bases (essayer de faire petit) BASE 1 (1 et 2 correspondent au MIS) 1 - 1 - 100 points - Average size : 35 points - Average time : 0.00129 s - Fails : 0 1 - 2 - 100 points - Average size : 37 points - Average time : 9.1E-4 s - Fails : 0 1 - 1 - 500 points - Average size : 169 points - Average time : 0.0046 s - Fails : 0 1 - 2 - 500 points - Average size : 174 points - Average time : 0.00425 s - Fails : 0 1 - 1 - 1000 points - Average size : 352 points - Average time : 0.01495 s - Fails : 0 1 - 2 - 1000 points - Average size : 364 points - Average time : 0.01403 s - Fails : 0 1 - 1 - 5000 points - Average size : 377 points - Average time : 0.29473 s - Fails : 0 1 - 2 - 5000 points - Average size : 394 points - Average time : 0.25969 s - Fails : 0 1 - 1 - 10000 points - Average size : 386 points - Average time : 1.16013 s - Fails : 0 1 - 2 - 10000 points - Average size : 402 points - Average time : 1.01815 s - Fails : 0

BASE 2 (1 et 2 correspondent au MIS) 2 - 1 - 100 points - Average size : 36 points - Average time : 2.0E-4 s - Fails : 0 2 - 2 - 100 points - Average size : 37 points - Average time : 2.4E-4 s - Fails : 0 2 - 1 - 500 points - Average size : 173 points - Average time : 0.00398 s - Fails : 0 2 - 2 - 500 points - Average size : 179 points - Average time : 0.00375 s - Fails : 0 2 - 1 - 1000 points - Average size : 351 points - Average time : 0.0149 s - Fails : 0 2 - 2 - 1000 points - Average size : 365 points - Average time : 0.01398 s - Fails : 0 2 - 1 - 5000 points - Average size : 377 points - Average time : 0.29543 s - Fails : 0 2 - 2 - 5000 points - Average size : 393 points - Average time : 0.26125 s - Fails : 0 2 - 1 - 10000 points - Average size : 384 points - Average time : 1.15217 s - Fails : 0 2 - 2 - 10000 points - Average size : 400 points - Average time : 1.01027 s - Fails : 0

BASE 3 (1 et 2 correspondent au MIS) 3 - 1 - 100 points - Average size : 16 points - Average time : 2.9E-4 s - Fails : 0 3 - 2 - 100 points - Average size : 18 points - Average time : 2.4E-4 s - Fails : 0 3 - 1 - 500 points - Average size : 172 points - Average time : 0.0039 s - Fails : 0 3 - 2 - 500 points - Average size : 180 points - Average time : 0.00367 s - Fails : 0 3 - 1 - 1000 points - Average size : 352 points - Average time : 0.01476 s - Fails : 0 3 - 2 - 1000 points - Average size : 365 points - Average time : 0.01395 s - Fails : 0 3 - 1 - 5000 points - Average size : 873 points - Average time : 0.32062 s - Fails : 0 3 - 2 - 5000 points

- Average size : 907 points - Average time : 0.28485 s - Fails : 0 3 - 1 - 10000
points - Average size : 1014 points - Average time : 1.20708 s - Fails : 0 3 - 2 -
10000 points - Average size : 1054 points - Average time : 1.0546 s - Fails : 0

3.2 Discussion

4 Conclusion

Références

- [1] Mihaela Cardei, Maggie Xiaoyan Cheng, Xiuzhen Cheng, and Ding-Zhu Du. Connected domination in multihop ad hoc wireless networks. In *JCIS*, pages 251–255, 2002.
- [2] Bo Gao, Yuhang Yang, and Huiye Ma. A new distributed approximation algorithm for constructing minimum connected dominating set in wireless ad hoc networks. *International Journal of Communication Systems*, 18(8) :743–762, 2005.
- [3] Yingshu Li, My T Thai, Feng Wang, Chih-Wei Yi, Peng-Jun Wan, and Ding-Zhu Du. On greedy construction of connected dominating sets in wireless networks. *Wireless Communications and Mobile Computing*, 5(8) :927–932, 2005.
- [4] Wikipedia. Graphe de disques — wikipedia, the free encyclopedia. https://fr.wikipedia.org/wiki/Graphe_de_disques. [consulté le 15-Oct-2017].
- [5] Wikipedia. Maximal independent set — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Maximal_independent_set. [consulté le 15-Oct-2017].