



# Évaluer et aménager les réseaux de transport

Thomas FERNIQUE<sup>1</sup>  
Rapport de stage de MIM 2

réalisé au COGIT<sup>2</sup> sous la direction de  
Jean-François Gleyze<sup>3</sup>

## Résumé

L'étude des réseaux de transport en géographie peut conduire à des problèmes combinatoires difficiles. L'algorithmique est susceptible d'aider à leur résolution, mais souffre d'un manque de formalisation en amont. Durant ce stage, j'ai cherché à relier géographie et informatique, en particulier pour obtenir des résultats sur l'efficacité et la vulnérabilité des réseaux de transport.

## Mots clés

algorithmique, dilatation, efficacité, géographie, graphes, vulnérabilité

---

1. thomas.fernique@ens-lyon.fr

2. Laboratoire COGIT (Recherche en bases de données géographiques) - Institut Géographique National,  
2/4 avenue Pasteur - 94165 Saint-Mandé Cedex

3. jean-francois.gleyze@ign.fr

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Concepts de base</b>	<b>2</b>
1.1 Modélisation des réseaux . . . . .	2
1.2 Efficacité . . . . .	3
1.3 Vulnérabilité . . . . .	4
<b>2 La dilatation</b>	<b>6</b>
2.1 Indices de Shimmel et de Kansky . . . . .	7
2.2 Dilatation . . . . .	7
2.3 Exemples sur quelques réseaux . . . . .	8
<b>3 Utilisation</b>	<b>11</b>
3.1 Différents objectifs . . . . .	11
3.2 Arbres couvrants, sous-graphes et réseaux efficaces . . . . .	12
3.3 Sous-réseaux renforcés . . . . .	14
<b>Conclusion</b>	<b>18</b>
<b>A Annexe</b>	<b>19</b>
A.1 Densité angulaire d'une surface . . . . .	19
A.2 Un résultat de NP-complétude . . . . .	20
A.3 Exemple de non-optimalité de l'algorithme glouton . . . . .	21
A.4 Minimiser l'augmentation de dilatation . . . . .	22
A.5 Indices de Shimmel et Kansky sur le réseau routier français . . . . .	27
<b>Références</b>	<b>29</b>

## Introduction

Le terme générique *réseau* désigne une structure organisée reliant d’une manière plus ou moins complexe des éléments, ou sommets, distincts. Ainsi les réseaux d’électricité ou d’eau vont relier les bâtiments ou villes, un réseau informatique relie des ordinateurs distants et dans un réseau social les gens pourront être considérés “reliés” lorsque par exemple ils se connaissent. Dans le cadre d’un réseau de transport, les éléments à relier seront les *usagers*, ceci grâce à diverses voies de communications : routes, voies ferrées, canaux *etc.*

Mais chaque type de réseau possède une logique qui lui est propre et doit être prise en compte pour une étude un tant soit peu poussée : la localisation spatiale des sommets importe peu pour les réseaux sociaux ou informatiques alors qu’elle est primordiale pour les réseaux de transport ou d’électricité/eau ; l’information se duplique mais ni l’électricité ni l’eau ; les usagers et l’information cherchent des déplacements rapides alors que l’électricité ou l’eau se répartissent sur le réseau selon les résistances ou engorgements de celui-ci . . . Dans ce rapport, on s’intéresse plus spécifiquement aux réseaux de transport. Avec quelques restrictions : ni les réseaux ni leur utilisation ne dépendront du temps (pas d’horaire de train ni de migration pendulaire), et les questions de congestion sont écartées. Ce choix n’exclut certes pas l’utilisation, plus ou moins directe, de résultats plus généraux ou bien propres à d’autres réseaux.

Les réseaux de transport sont au cœur des problèmes d’aménagement du territoire. En effet, leur efficacité plus ou moins grande à desservir et relier entre eux les différents points du territoire va conditionner le développement de celui-ci. Cette efficacité est par ailleurs susceptible de diminuer suite à des endommagements des réseaux : la vulnérabilité de ceux-ci s’avère donc aussi un facteur important. Construction ou amélioration des réseaux de transport doivent donc se faire dans une optique d’efficacité maximale et vulnérabilité minimale. Ceci requiert notamment de savoir évaluer ces notions, c’est-à-dire de les définir clairement et de se donner les moyens de les mesurer.

La construction de réseaux de transport efficaces ou peu vulnérables est un problème peu traité en informatique théorique car les concepts d’efficacité ou de vulnérabilité sont inévitablement subjectifs. Pourtant, en supposant définis formellement ces concepts, l’apport de l’informatique pour élaborer des algorithmes tant d’évaluation que de construction pourrait être très profitable. Or ces concepts de base sont souvent le sujet même d’étude d’autres domaines scientifiques, notamment des sciences géographiques dans le cadre des réseaux de transport. Aussi pourrait-il être productif, et c’est ce que ce rapport tente de montrer, de travailler à cheval sur ces deux tableaux. D’autant plus que la séparation entre recherche en informatique théorique et en géographie, malgré les besoins de cette dernière en algorithmique, est souvent assez grande, ce qui conduit parfois à la redécouverte plus ou moins heureuse de résultats déjà établis.

Ceci dit, si beaucoup d’articles de géographie s’intéressent à l’évaluation des réseaux, ceux proposant une mesure formelle sont rares. Surtout, je n’ai trouvé qu’un seul article, [BR97], cherchant à déterminer ce que devrait vérifier *toute* mesure (ceci dans un contexte de mesure de vulnérabilité de réseaux informatique). Il propose pourtant une alternative intéressante au parachutage d’une formule puis à sa justification sur quelques exemples. On s’inspirera de cette démarche pour ébaucher une réflexion, même si elle relève plutôt de compétences en géographie qu’en informatique, sur les critères d’une mesure d’efficacité des réseaux de transport.

Ce rapport est articulé en trois parties : réflexion sur les concepts d’efficacité et de vulnérabilité ; introduction d’une mesure d’efficacité particulière, comparaison avec d’autres mesures et exemples d’évaluation de réseaux ; utilisation de la mesure précédente pour construire des réseaux : en particulier on s’intéressera, pour un réseau donné, au problème de détermination d’un sous-réseau efficace à protéger en priorité.

# 1 Concepts de base

## 1.1 Modélisation des réseaux

Le territoire étudié est assimilé à une partie  $\mathcal{T}$  de  $\mathbb{R}^2$ .

Un réseau sur  $\mathcal{T}$  est modélisé par un *graphe pondéré*  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, t, c)$  où :

- $\mathcal{V} = \{S_1, S_2, \dots, S_n\}$  est l'ensemble des *sommets* ou *points d'accès* (gares, carrefours ...);
- $\mathcal{E} = \{e_1, e_2, \dots, e_m\} \subset \mathcal{V} \times \mathcal{V}$  est l'ensemble des arêtes (routes, voies ferrées ...);
- $t: \mathcal{E} \mapsto \mathbb{R}^+$  associe à une arête son temps de parcours;
- $c: \mathcal{E} \mapsto \mathbb{R}^+$  associe à une arête son coût (longueur par exemple).

On suppose en outre connue la *matrice origines-destinations*  $\Phi$ , où le coefficient  $\phi_{ij}$  est la proportion de déplacements sur le réseau se faisant du point  $S_i$  au point  $S_j$ .

Notons que les déplacements ne se font *a priori* pas uniquement sur le réseau. Si par exemple le réseau considéré est un réseau routier, les usagers peuvent aussi se déplacer à pieds pour accéder au réseau. De plus, étant donnée la structure généralement *multi-échelles* des réseaux de transport (le réseau autoroutier a une structure similaire au réseau des routes nationales mais à plus grande échelle), il arrive souvent que le réseau auquel on s'intéresse ne soit qu'une *restriction* du réseau réel à ses axes les plus importants : il est alors comme dessiné sur un réseau "de fond" sur lequel peuvent aussi s'effectuer les déplacements. On considérera très simplement que ces déplacements se font à une vitesse constante  $v_{\mathcal{T}}$  caractéristique de ce réseau "de fond". Si par exemple on s'intéresse au réseau routier sans les routes communales, on peut modéliser grossièrement celles-ci en considérant qu'elle permettent un déplacement à la vitesse  $v_{\mathcal{T}} = 30\text{km/h}$  en dehors du réseau.

On distingue deux "distances" :

- Une mesure  $d_{\mathcal{G}}$  sur le réseau :  $d_{\mathcal{G}}(S_i, S_j)$  est le plus petit temps de parcours sur le réseau<sup>4</sup> entre deux sommets;
- Une mesure de *référence*  $d^*$  :  $d^*(P, Q)$  est le temps de parcours idéal entre deux points quelconques du territoire. On prendra généralement le temps de parcours en ligne droite à la vitesse maximale possible sur le réseau étudié.

Enfin, on modélise le temps d'accès au réseau. Les usagers sont supposés se rendre au point d'accès du réseau le plus proche d'eux (ce qui n'est pas forcément le cas : on peut par exemple préférer marcher un peu plus pour prendre le métro à une station plus commode pour le déplacement envisagé), aussi chaque sommet  $S_i$  "couvrira" la zone  $\mathcal{V}_i$  du territoire constituée des points dont  $S_i$  est le plus proche sommet<sup>5</sup>. On définit alors les quantités :

$$\pi_i = \int_{\mathcal{V}_i} d_{\mathcal{T}}(P, S_i) dP$$

$$\pi_i^* = \int_{\mathcal{V}_i} d^*(P, S_i) dP$$

Où  $d_{\mathcal{T}}$  mesure le temps de parcours en ligne droite à vitesse  $v_{\mathcal{T}}$ . Ainsi  $\pi_i$  (resp.  $\pi_i^*$ ) est le temps moyen mis par les usagers de la zone couverte par  $S_i$  pour joindre  $S_i$  par le réseau "de fond" (resp. par un réseau idéal).

---

4. éventuellement on permet des déplacements sur le territoire à la vitesse  $v_{\mathcal{T}}$  (il peut être plus rapide, dans le cas du métro parisien, de relier à pieds deux stations de métro !): ceci revient à ajouter une arête de poids *ad hoc* entre toute paire de sommets non reliés.

5.  $\mathcal{V}_i$  est le polygône de Voronoï associé à  $S_i$

## 1.2 Efficacité

### Pourquoi mesurer l'efficacité

Définir - mathématiquement - une mesure d'efficacité sur les réseaux de transport doit permettre de les comparer, de déterminer leurs points faibles et de les améliorer ou optimiser. En fait, on peut ainsi assujettir l'étude des réseaux à un cadre formel permettant d'adopter une démarche rigoureuse - et on l'espère productive - où la validité des résultats ne dépendra que du choix initial de la mesure. Autrement dit, on cherche à définir un cadre subjectif satisfaisant dans lequel on pourra développer une théorie objective consistante.

### Définir l'efficacité

Efficace : qui produit l'effet qu'on en attend (Petit Robert).

Comment formaliser cette définition dans le cadre des réseaux de transport? Il faudrait définir ce qu'on attend d'un réseau de transport et quantifier dans quelle mesure un réseau donné produit l'effet escompté, c'est-à-dire répond aux attentes. Intuitivement, ce qu'on attend d'un réseau de transport est de *faciliter* le déplacement des usagers sur le territoire. Nous conviendrons ici qu'un déplacement est d'autant plus facile qu'il est rapide. Une étude plus sociologique des attentes des usagers conduirait sans doute à une définition plus nuancée, mais celle-ci semble assez réaliste pour une première approche.

Il faut aussi distinguer deux composantes de l'efficacité. La première, *intrinsèque* au réseau, ne dépendant que de la structure de celui-ci : le déplacement sur le réseau. La deuxième, liée à la *desserte* du territoire, dépend de l'implantation (densité et répartition) des sommets : l'accès au réseau (Figure 1).

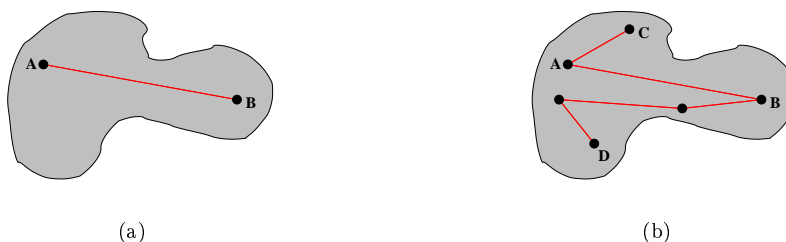


FIG. 1 – En (a) l'efficacité intrinsèque est bonne (A et B sont faciles à relier) mais la desserte mauvaise. En (b) la desserte est meilleure, mais le réseau relie mal C et D.

Quel serait dans ce cadre un réseau idéal? La desserte est optimale pour une densité infinie de sommets, et l'efficacité intrinsèque pour un réseau complet, c'est-à-dire où tous les sommets sont reliés deux à deux. Malheureusement c'est irréaliste : pour le réseau routier français il faudrait transformer le pays en un gigantesque parking! L'objectif sera donc plutôt de trouver des réseaux *bons* ou *meilleurs*, ceci grâce à une mesure quantifiant le mieux possible l'efficacité.

### Des critères pour mesurer

Avant de définir une mesure d'efficacité particulière, il peut être intéressant de définir ce que devrait vérifier *toute* mesure d'efficacité. Cette généralisation centrerait vraiment la réflexion sur le concept d'efficacité. Voici quelques critères retenus pour une mesure d'efficacité (tout en souhaitant que des personnes plus qualifiées dans ce domaine poursuivent cet essai) :

1. Elle doit être à valeur dans un ensemble totalement ordonné. Ainsi on peut comparer entre eux deux réseaux desservant un même territoire, donner un sens à la notion de

*réseau optimal* et chercher comment améliorer (ou détériorer !) un réseau.

2. Elle ne peut qu'augmenter si on ajoute un élément (arête ou nœud) au réseau. En effet l'utilisateur peut toujours ignorer les ajouts, auquel cas le réseau lui rendra les mêmes services et sera donc au moins aussi efficace. Inversement, l'efficacité ne peut que diminuer si on supprime un élément du réseau car, l'utilisateur étant supposé l'utiliser au mieux, il ne pourra qu'être gêné (et non découvrir de meilleurs itinéraires).
3. Elle doit tenir compte du fait que l'efficacité n'est pas une notion locale : une amélioration ponctuelle peut très bien entraîner une détérioration à l'échelle du réseau entier. Ainsi figure 2, le carré central est moins efficace en (a) qu'en (b) (même réseau symétrisé mais avec une arête en plus) ; pourtant la perte d'efficacité en (b) due à l'allongement des liaisons entre  $G_1$  et  $G_2$  peut, à l'échelle du réseau entier, effacer le gain dû à l'amélioration (locale) du carré central.
4. Elle doit être indépendante de la taille ou de la forme du territoire desservi : c'est le réseau qui est évalué et non le territoire !
5. Éventuellement, elle permet de localiser les points faibles du réseau, par exemple en mesurant l'efficacité d'une sous-région relativement au réseau entier (desserte locale et facilité à relier la sous-région au reste du réseau).

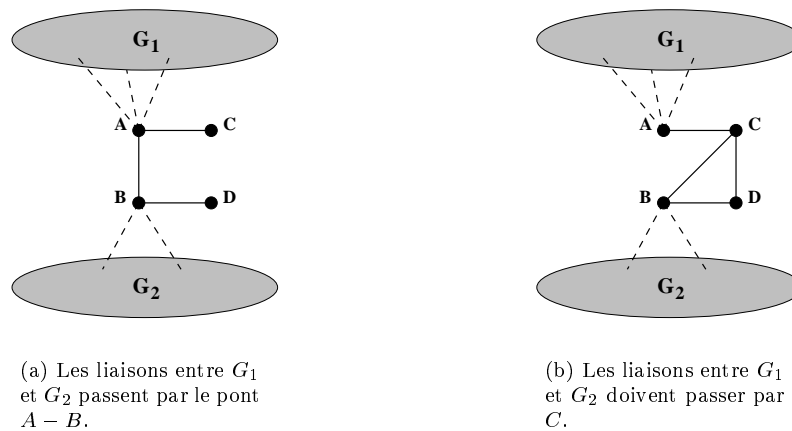


FIG. 2 – Un exemple d'amélioration locale mais non globale.

Les deux premières conditions me semblent nécessaires. Les deux suivantes, plus intuitives, ont pour but d'éviter certains écueils plutôt que d'aider à construire une mesure. La dernière peut être utile pour étudier les possibilités d'amélioration/détérioration du réseau.

Voyons à présent comment le concept d'efficacité permet de définir celui de vulnérabilité.

### 1.3 Vulnérabilité

#### Vulnérabilité et baisse d'efficacité

**Vulnérable** : qui peut être facilement atteint, se défend mal (Petit Robert).

Dans le cas d'un réseau de transport, "atteint" signifie endommagé. Or il semble logique de considérer qu'un réseau subit un dommage - à travers la rupture d'une arête ou d'un sommet - si et seulement s'il devient moins efficace. Un réseau serait donc d'autant plus vulnérable que son efficacité peut être facilement diminuée. Ainsi on peut déduire d'une mesure d'efficacité donnée une mesure de vulnérabilité. Il reste cependant une difficulté : la baisse d'efficacité (donc la vulnérabilité) dépend de l'endommagement du réseau alors qu'on voudrait une mesure synthétique intégrant les diverses avaries envisageables.

### Fonction de vulnérabilité d'un réseau

Si on quantifie un dommage par le *coût de la reconstruction* qu'il entraîne, la vulnérabilité d'un réseau peut être représentée par une courbe donnant la baisse d'efficacité en fonction de ce coût de reconstruction. Pourtant à coût fixé, la baisse d'efficacité n'est pas bien définie : en effet des dommages de même coût peuvent affecter différemment le réseau selon leur localisation. Aussi faut-il faire un choix : on peut prendre une baisse d'efficacité *moyenne* sur tous les dommages de même coût, ou bien prendre le maximum. . . Dans le cadre de dommages naturels, une moyenne semble plus adaptée (avec éventuellement pour une modélisation plus fine une pondération géographique des dommages : telle route en bord de fleuve a plus de chance d'être inondée que celle-ci en haute montagne). On obtient alors des courbes du type de celle figure 3. De manière générale, la courbe croît avec le coût de reconstruction et atteint un maximum pour un réseau rendu totalement inefficace. Des propriétés plus précises (continuité, seuils . . .) dépendent *a priori* du réseau étudié.

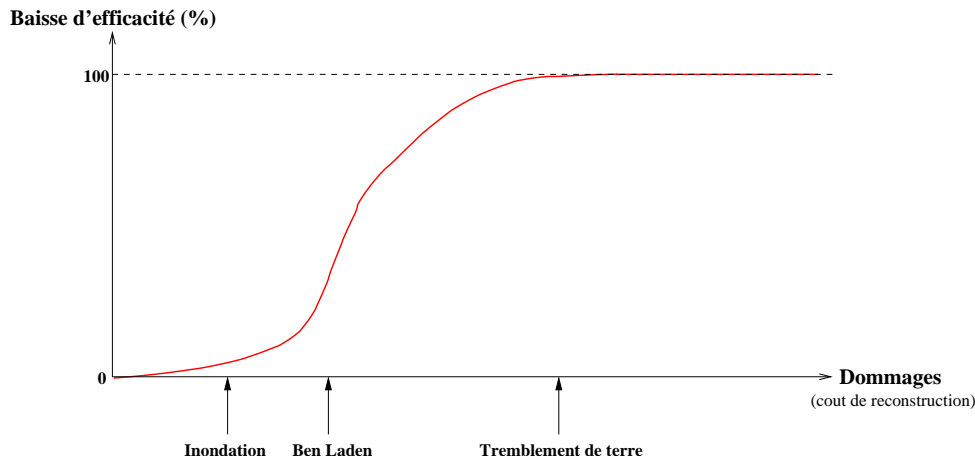


FIG. 3 – *Fonction de vulnérabilité d'un réseau*

### Comparaisons de vulnérabilités

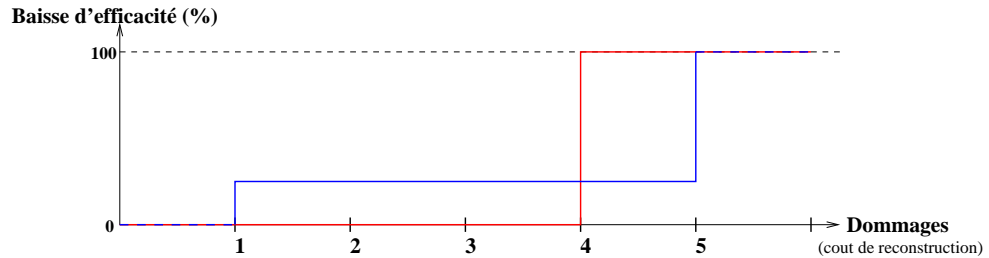
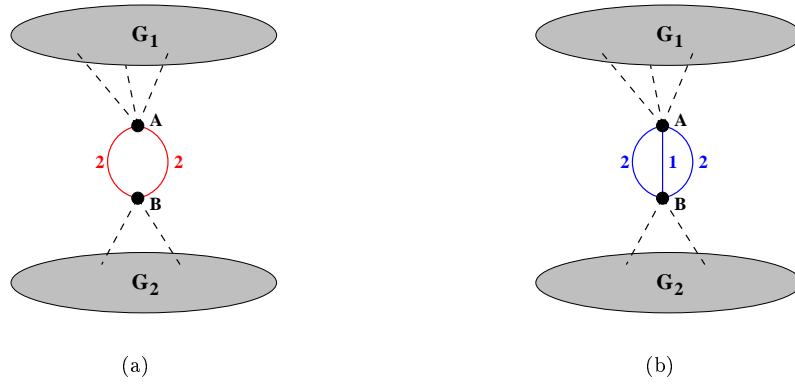
Le défaut d'une représentation de la vulnérabilité par une courbe est qu'on ne peut pas toujours comparer deux fonctions. Que dire quand deux fonctions de vulnérabilité se croisent ? On peut en effet très bien concevoir deux réseaux tels que l'un soit moins vulnérable que l'autre pour de petits dommages, mais plus pour des dommages très importants.

La figure 4 illustre ceci. On se restreint à des dommages sur les arêtes entre les sommets  $A$  et  $B$ . En 4(a), la rupture d'une seule arête (coût 2) n'affecte pas l'efficacité tandis que la rupture de deux arêtes (coût 4) déconnecte le réseau et affecte ainsi fortement son efficacité. En 4(b), la rupture de l'arête centrale (coût 1) affecte l'efficacité puisque les liaisons entre  $\mathcal{G}_1$  et  $\mathcal{G}_2$  doivent se faire par des arêtes plus longues. Il faut alors rompre les deux arêtes restantes (coût 5 au total) pour déconnecter le réseau. Les courbes de vulnérabilité de ces réseaux sont représentées figure 4(c) : le réseau le plus vulnérable change selon l'importance des dommages.

### Scenarii d'endommagement et courbe de Farmer

On vient de voir qu'il est problématique de comparer des réseaux parce que leur vulnérabilité dépend de l'importance des dommages. On propose ici deux solutions envisageables.

On peut n'avoir besoin d'estimer la vulnérabilité de réseaux que face à des circonstances assez précises : on cherchera par exemple à construire un réseau routier le moins vulnérable



(c) En rouge (resp. bleu) la courbe de vulnérabilité du réseau  $a$  (resp.  $b$ )

FIG. 4 – Deux réseaux et leurs courbes de vulnérabilité

possible aux inondations, sans se préoccuper des pannes de feux tricolores ou des attaques thermonucléaires. Dans ce cas, plutôt que de déterminer la courbe de vulnérabilité entière, on calcule les baisses d'efficacité consécutives à un scénario d'endommagement précis : on peut par exemple considérer la rupture de routes localisées en bord de fleuve pour étudier la résistance du réseau aux inondations.

Si par contre on veut une mesure de vulnérabilité synthétisant aussi bien que possible de la résistance du réseau aux différents dommages qu'il peut subir, une moyenne pondérée semble offrir une solution acceptable. À cette fin, on utilise ce qui est connu sous le nom de *courbe de Farmer*, à savoir la courbe associant à un dommage sa fréquence (ou probabilité) d'arriver. Dans le cadre de notre modélisation, on associera à l'ensemble des dommages de coût  $x$  une probabilité d'occurrence  $p_r(x)$  : on obtient une courbe comme celle figure 5. En notant  $\Delta(x)$  la baisse d'efficacité consécutive aux dommages de coût  $x$ , on définit alors la vulnérabilité par :

$$\int_0^{\infty} \Delta(x)p_r(x)dx$$

## 2 La dilatation

Ici nous introduirons le concept de *dilatation* pour mesurer l'efficacité des réseaux. On commencera tout d'abord par donner les deux seuls indices déjà existant pour mesurer l'efficacité, avant de montrer quels avantages apporte la dilatation. La section 2.3 donne quelques exemples de calcul de dilatation afin de se familiariser avec cette notion.



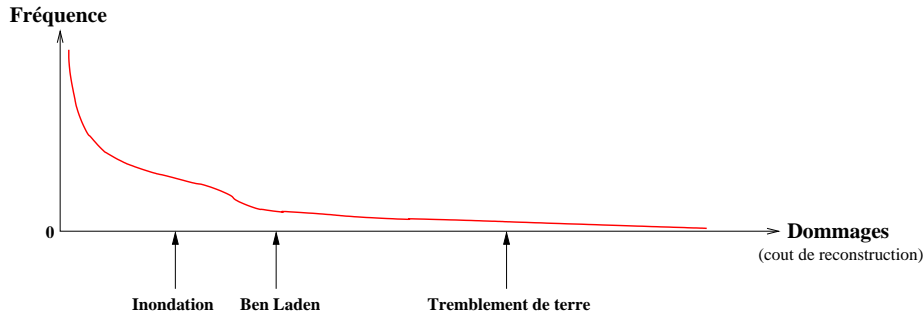


FIG. 5 – Fréquence d'occurrence des différents dommages

## 2.1 Indices de Shimbél et de Kansky

Les deux indices présentés ci-dessous sont les seuls trouvés pouvant être vus comme des mesures d'efficacité, bien que ce lien ne semble jamais avoir été explicitement fait, ni dans les articles d'origines, ni dans ceux citant ces derniers.

Pour deux sommets  $S_i$  et  $S_j$ , on pose  $\sigma_{ij} = d_G(S_i, S_j)$ . On note  $\sigma_i$  la moyenne à  $i$  fixé des  $\sigma_{ij}$  et  $\sigma$  la moyenne de tous les  $\sigma_{ij}$ .  $\sigma_i$  est l'*accessibilité relative* du sommet  $S_i$ , et  $\sigma$  l'*indice d'accessibilité* ou *indice de Shimbél* du réseau ([Shi53]).

À partir de  $\kappa_{ij} = (d_G(S_i, S_j) - d(S_i, S_j))^2$ , détour entre  $S_i$  et  $S_j$ , on définit de la même manière  $\kappa_i$  et  $\kappa$ .  $\kappa$  est l'*indice de Kansky* du réseau ([Kan89]).

Un réseau est alors jugé d'autant plus efficace qu'il a un faible indice. Ces deux indices présentent cependant un défaut gênant : ils dépendent beaucoup du territoire. Dans le cas du réseau routier français par exemple, l'indice d'accessibilité de la ville de Brest sera naturellement élevé du fait de la position excentrée de la ville et donc de son implication dans des liaisons longues en moyenne ; ceci même si de nombreuses autoroutes relient la ville aux autres villes. L'indice de Kansky accordera la même valeur à une route presque rectiligne de 850km reliant deux villes distantes (à vol d'oiseau) de 800km et à une route sinueuse de 100km entre deux villes distantes de 50km, ce qui peut être critiqué. On constate aussi que plus le territoire est grand, plus l'efficacité a tendance à être faible, même pour une structure de réseau (forme et densité) équivalente (Figure 6). La condition 4 présentée section 1.2 n'est donc pas vérifiée.

## 2.2 Dilatation

On définit la *dilatation* entre deux sommets  $S_i$  et  $S_j$  par :

$$\delta_{ij} = \frac{d_G(S_i, S_j)}{d^*(S_i, S_j)}$$

Dans la suite de ce rapport, on se limitera à étudier l'efficacité intrinsèque au réseau. Ceci correspond notamment au cas où les sommets sont des villes qui concentrent *tous* les usagers du réseau. La prise en compte de la desserte peut se faire soit en parallèle, soit en remplaçant  $\delta_{ij}$  dans tout ce qui suit par une dilatation  $\tilde{\delta}_{ij}$  avec pénalités (définies section 1.1) définie par :

$$\tilde{\delta}_{ij} = \frac{\pi_i + d_G(S_i, S_j) + \pi_j}{\pi_i^* + d(S_i, S_j) + \pi_j^*}$$

Comme pour les indices précédents, on définit la dilatation  $\delta_i$  relative au sommet  $S_i$  et la *dilatation moyenne*  $\delta$ . On considère alors qu'un réseau est d'autant plus efficace que sa dilatation moyenne est faible, c'est-à-dire proche de 1 (cas du réseau idéal, infiniement dense et reliant tous les couples de sommets deux à deux). L'intérêt par rapport aux indices

précédents est de s'affranchir de cette dépendance vis-à-vis du territoire. La figure 6 montre comment évoluent les deux indices précédents et la dilatation sur un réseau simple : une grille régulière de  $n \times n$  mailles carrées. Quand la grille grandit, les indices de Shimbel et Kansky augmentent, et l'efficacité associée tend vers 0. Par contre la dilatation se stabilise (elle tend vers  $\frac{4}{\pi}$  précisément) et mesure une efficacité propre à la structure du réseau. Par ailleurs, dans le cas d'une ville excentrée sur le territoire, un réseau sera jugé efficace s'il la relie bien, relativement à ce qu'on peut attendre au vu de sa position géographique, au reste du réseau : c'est en pratique le cas de villes comme Lille ou Marseille dans le cadre du réseau routier français.

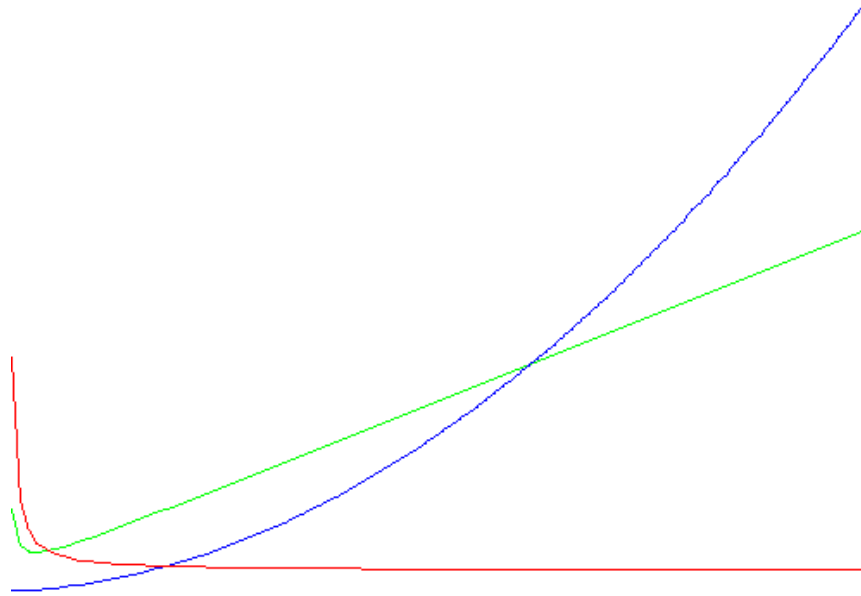


FIG. 6 – Variations de l'indice de Shimbel (vert), de l'indice de Kansky (Bleu) et de la dilatation (rouge) en fonction du nombre  $n^2$  de sommets d'un réseau constitué d'une grille carrée régulière  $n \times n$  (style Manhattan).

La condition 4 (section 1.2) semble donc vérifiée. Les autres conditions aussi, sauf le cas de l'ajout d'un sommet : en effet dans le cadre d'une modélisation où l'utilisateur se rend au sommet le plus proche de lui pour accéder au réseau (et non au plus avantageux pour le déplacement qu'il envisage), un nouveau sommet peut drainer des usagers mais rallonger leur déplacement. Il faudrait pour remédier à cela prendre un modèle plus fin du choix du sommet d'accès au réseau pour un usager.

## 2.3 Exemples sur quelques réseaux

### Réseaux à mailles régulières

Les réseaux réguliers à mailles carrées ou triangulaires présentent une structure proche de certains réseaux réels (réseaux routiers urbains par exemple). Malgré leur simplicité, établir une formule de la dilatation moyenne s'avère rapidement impraticable. Aussi s'intéressera-t-on plutôt au comportement asymptotique, c'est-à-dire pour une densité de sommets ou une taille de réseau qui tend vers l'infini.

On remarque pour ces réseaux que la dilatation entre deux points  $A$  et  $B$  ne dépend que de la pente  $\alpha$  de la droite  $AB$  : on notera  $\delta(\alpha)$  cette dilatation. Pour des mailles carrées :  $\delta(\alpha) = \cos \alpha + \sin \alpha$  ; pour des mailles triangulaires (équilatérales) :  $\delta(\alpha) = \cos \alpha + \frac{1}{\sqrt{3}} \sin \alpha$ . Si on connaît alors, pour tout  $\alpha$ , la proportion  $p(\alpha)$  de couples de sommets qui forment une droite de pente  $\alpha$ , on peut calculer la dilatation moyenne par la formule :  $\delta = \int_0^\pi \delta(\alpha)p(\alpha)d\alpha$ .

En tournant le réseau d'un angle  $\alpha_0$  (le territoire reste fixe), cette formule devient :  $\delta = \int_0^\pi \delta(\alpha) p(\alpha - \alpha_0) d\alpha$ .

Pour un réseau infiniment dense, la fonction  $p$  est la densité de probabilité de la pente d'un segment tiré au hasard sur le territoire. Elle dépend donc uniquement du territoire. Pour un disque par exemple, la densité sera uniforme (pentes équiprobables) valant  $\frac{1}{\pi}$ . Pour d'autres surfaces, le calcul est généralement plus difficile : voir en annexe page 19 pour plus de détails. Dans tous les cas :

$$\frac{1}{\pi} \int_0^\pi \int_0^\pi \delta(\alpha) p(\alpha - \alpha_0) d\alpha d\alpha_0 = \frac{1}{\pi} \int_0^\pi \delta(\alpha) \underbrace{\int_0^\pi p(\alpha - \alpha_0) d\alpha_0}_{=1} d\alpha = \frac{1}{\pi} \int_0^\pi \delta(\alpha) d\alpha$$

Et cette quantité ne dépend plus de la forme du territoire mais seulement du type de maillage. Elle vaut par exemple  $\frac{4}{\pi} \simeq 1.27$  pour des mailles carrées et  $\frac{2\sqrt{3}}{\pi} \simeq 1.10$  pour des mailles triangulaires. De plus, par continuité et positivité de  $\alpha_0 \mapsto \int_0^\pi \delta(\alpha) p(\alpha - \alpha_0) d\alpha$  il existe  $\alpha_0^*$  tel que  $\frac{1}{\pi} \int_0^\pi \delta(\alpha) p(\alpha - \alpha_0^*) d\alpha \leq \frac{1}{\pi} \int_0^\pi \delta(\alpha) d\alpha$ . En clair : quelque soit la forme du territoire, on peut orienter (angle  $\alpha_0^*$ ) un réseau à mailles carrées (resp. triangulaires) infiniment dense sur ce territoire pour obtenir dilatation moyenne inférieure à  $\frac{4}{\pi}$  (resp.  $\frac{2\sqrt{3}}{\pi}$ ). Dans le cas plus réaliste de réseaux finis, on peut approcher arbitrairement près cette dilatation et même l'atteindre si le territoire n'est pas isotrope (densité  $p$  non constante).

Notons par ailleurs que les mailles triangulaires offrent une dilatation maximale faible :  $\frac{2}{\sqrt{3}} \simeq 1.15$  ( $\sqrt{2} \simeq 1.41$  pour les mailles carrées).

### Réseau de Delaunay

Les réseaux réels ont cependant rarement une structure aussi régulière que les réseaux précédents. On s'intéresse ici à une structure plus souple : la triangulation de Delaunay (voir [BY95] pour la définition et la construction). En fait on s'aperçoit que beaucoup de réseaux réels ont une structure très proche.

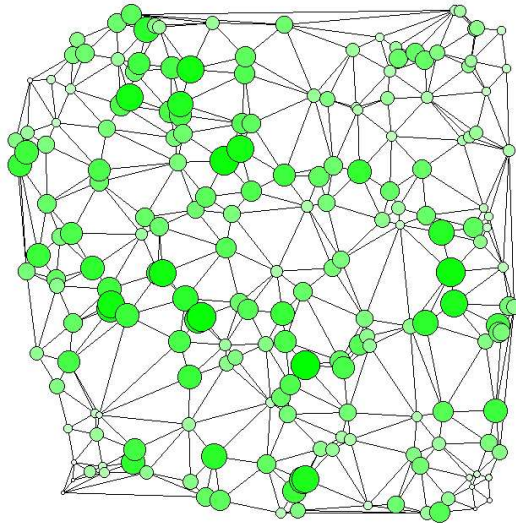


FIG. 7 – Dilatation sur un réseau de Delaunay de 200 points. La taille d'un disque est proportionnelle à la dilatation relative du sommet. La dilatation moyenne vaut 1.0539

Quelques résultats théoriques ont été démontrés sur la dilatation de tels réseaux. Dans [KG89], il est montré que la dilatation maximale entre deux sommets d'un réseau de Delaunay est inférieure à  $\frac{2\pi}{3 \cos \frac{\pi}{6}} \simeq 2.42$  (le fait même que la dilatation soit bornée n'est pas

trivial). Cette majoration semble pourtant large : si Paul Chew montre dans [Che89] qu'on peut atteindre des dilatations arbitrairement proches de  $\frac{\pi}{2} \simeq 1.57$ , il n'a pas été trouvé de cas de dilatation supérieure. Aussi peut-on espérer une borne supérieure plus petite (peut-être  $\frac{\pi}{2}$  elle-même?). De plus la dilatation moyenne (qui est celle qui nous intéresse) est en général beaucoup plus faible : des calculs numériques sur des réseaux de tailles variées donnent des dilatations moyennes oscillant entre 1.05 et 1.10<sup>6</sup>, soit très proche du cas idéal du réseau complet (dilatation 1). Dans [BTZ98], il est démontré que sur un semis de sommets donné par un processus de Poisson (sommets répartis de manière homogène), la dilatation entre deux points est asymptotiquement (c'est-à-dire quand la distance entre les deux points tend vers l'infini) inférieure à  $\frac{4}{\pi} \simeq 1.27$ . Ainsi pour un grand réseau, la dilatation moyenne aura toutes les chances d'être inférieure à 1.27.

En conclusion, un réseau de Delaunay a en général une dilatation moyenne faible et constitue une bonne approximation du réseau complet, d'autant plus qu'il est planaire et comporte relativement peu d'arêtes (ce qui le rend envisageable en pratique).

### Métro parisien

L'exemple du métro parisien est un peu particulier car le réseau n'est pas vraiment constitué d'arcs mais de *lignes* : une modélisation un peu fine doit prendre en compte les temps de correspondance. Ceci peut se faire par exemple en décomposant chaque station en autant de sommets qu'il y a de lignes qui se croisent et en reliant ces sommets entre eux par des arcs dont la longueur reflètent les temps de correspondances. Dans le cadre de ce simple exemple sur la dilatation on se contente d'un modèle très simplifié à temps de correspondance nuls (voir [Gle03] pour un modèle plus fin).

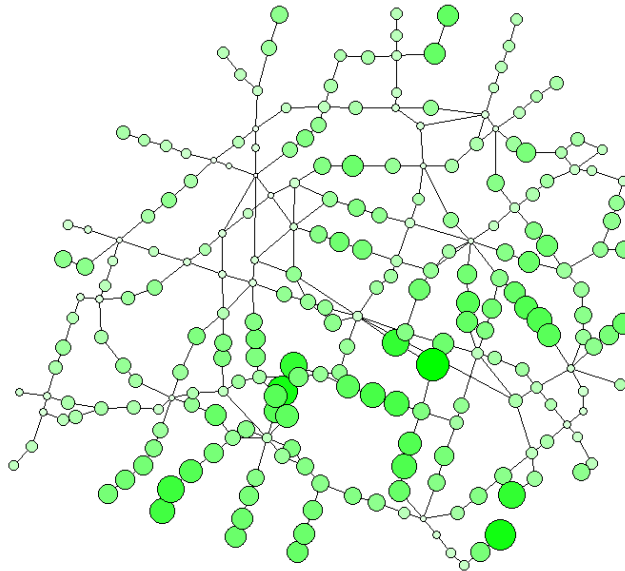


FIG. 8 – Dilatation sur le réseau métropolitain de Paris intra-muros. La dilatation moyenne vaut 1.25

### Réseau routier français

Le dernier exemple est celui du réseau routier français restreint aux plus grandes routes (figure 9 page 11). La distance de référence  $d^*$  entre deux sommets  $S_i$  et  $S_j$  est le temps de

6. On a d'ailleurs vu au paragraphe précédent qu'on pouvait approcher 1.10 en densifiant suffisamment un réseau de triangles équilatéraux. Or la triangulation de Delaunay a justement pour propriété d'être formée de triangles "compacts" au sens de "presque équilatéraux".

parcours à vol d'oiseau et à la vitesse de conduite sur autoroute.

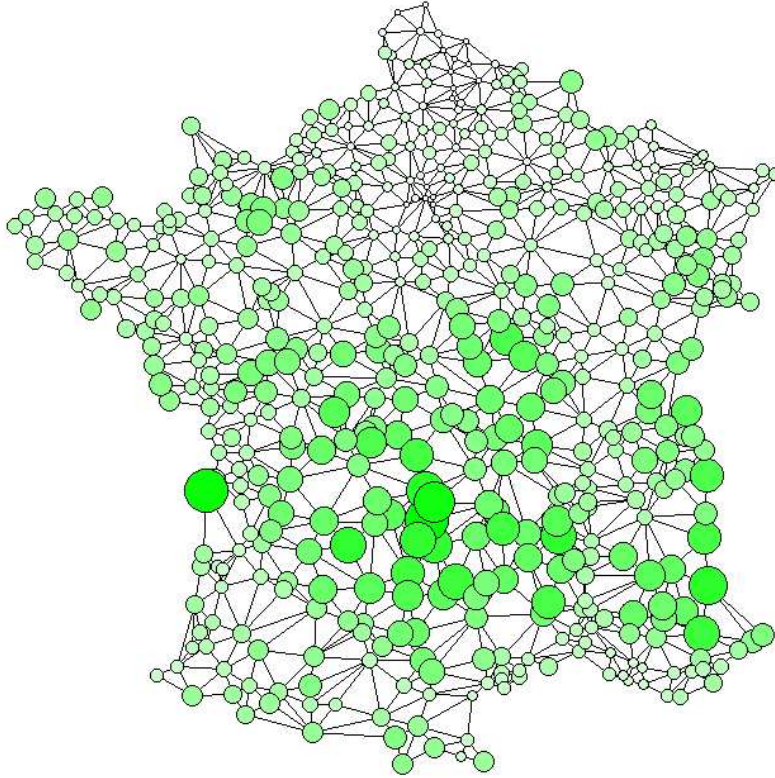


FIG. 9 – Dilatation sur le réseau routier français. La dilatation moyenne vaut 1.44

Visuellement, la structure est proche d'un réseau de Delaunay. Il pourrait d'ailleurs être intéressant de mesurer "l'écart" (notion à préciser) entre un tel graphe et une triangulation de Delaunay. Un commentaire exhaustif serait fastidieux, mais la dilatation fait incontestablement ressortir les éléments structurels du réseau : axes rapides (le couloir rhodanien est le plus flagrant) ou régions mal desservies (Massif Central bien sûr, mais les Vosges sont aussi visibles). Il semble plus difficile de tirer ces informations des indices de Shimmel et Kansky sur le même réseau (figures 21 et 22 page 27)...

## 3 Utilisation

### 3.1 Différents objectifs

Comment utiliser une mesure d'efficacité et la mesure de vulnérabilité associée à l'étude de réseaux réels? On peut par exemple évaluer globalement des réseaux et comparer nombre d'entre eux. On peut aussi, dans une logique de cartographie, utiliser des indices relatifs à un sommet ou une arête (on a par exemple vu la définition propre à un sommet des indices de Shimmel et de Kansky et de la dilatation) pour localiser et représenter des zones critiques : zones de faible efficacité handicapant le réseau ou zones dont l'endommagement provoque des dégâts importants à une échelle plus globale. Les figures 7, 8 ou 9 sont un exemple de ce qui est réalisable dans cette voie.

Dans ce rapport, nous avons plutôt utilisé cette approche pour *justifier* la validité de la mesure choisie. Le fait que la figure 9 mette en évidence un couloir rhodanien efficace est rassurant (pour les skieurs parisiens) mais peu intéressant : ce n'est pas la dilatation qui *révèle* l'existence de l'A7, n'importe quelle carte Michelin le fait bien mieux! Ceci est bien

sûr un exemple caricatural : en pratique certains travers ou particularités des réseaux sont délicats à mettre en évidence. Ainsi dans [Gle03], la cartographie de différents indicateurs sur le réseau du métro parisien a révélé des faits parfois inattendus : rôle prépondérant de stations que l'intuition juge "secondaires", mise en évidence de lignes "mal orientées" (écart important entre la direction globale du déplacement des usagers et celle de la ligne) . . . Mais une telle étude relève plus de compétences en géographie qu'en informatique, aussi avons nous cherché à utiliser la dilatation pour créer ou modifier des réseaux plutôt que pour interpréter ses résultats et réfléchir à ce qu'elle met en évidence dans tel réseau réel.

On s'intéresse notamment aux problèmes suivants :

- Construire un réseau *ex-nihilo* : si on se donne un semis de sommets sur lequel bâtir un réseau (choisir les positions de ces sommets est un problème différent), on veut déterminer un ensemble d'arcs relativement petit (coût de construction du réseau raisonable) conduisant à un réseau efficace ;
- Protéger un sous-réseau efficace : à défaut de pouvoir garantir un réseau complet invulnérable, on peut concentrer les efforts de protection, ou les priorités de reconstruction, sur un sous-réseau aussi efficace que possible. Un tel sous-réseau opérationnel assure une efficacité minimale du réseau complet ;
- Modifier un réseau : il peut s'agir d'étendre ce réseau (ajout d'une ligne de métro, construction d'une autoroute . . .) ou de l'améliorer localement (protection ou remodelage).

Notons au passage que le problème de construction d'un réseau *ex nihilo* se ramène à celui de la recherche de bons sous-réseau : on peut partir du réseau complet (tous les sommets sont reliés) et en déterminer un sous-réseau planaire efficace. Ces problèmes ont malheureusement été peu étudiés, surtout dans le contexte des réseaux de transport. La section suivante présente un état des recherches en informatique autour de ce sujet.

### 3.2 Arbres couvrants, sous-graphes et réseaux efficaces

On appelle *Minimal Communication Cost Spanning Tree (MCCST)* le problème suivant :

Étant donné un graphe  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  pondéré par  $w : \mathcal{E} \mapsto \mathbb{R}_*^+$  et une fonction de requête  $r : \mathcal{V}^2 \mapsto \mathbb{R}_*^+$ , on cherche un arbre couvrant  $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ , sous-graphe de  $\mathcal{G}$ , minimisant l'évaluation :

$$\sum_{\substack{u \neq v \\ u, v \in \mathcal{V}}} d_{\mathcal{G}'}(u, v) r(u, v)$$

Le lien avec la dilatation se fait en posant<sup>7</sup>  $r(u, v) = \frac{1}{n(n-1)} \frac{1}{d^*(u, v)} \Phi_{uv}$  : une solution optimale de *MCCST* est le plus efficace (au sens de la dilatation moyenne) des arbres sous-réseaux du réseau initial.

Malheureusement, il est démontré que le problème de décision associé à *MCCST* (existe-t-il un arbre d'évaluation inférieure à  $\delta$ ?) est NP-complet<sup>8</sup>. Il est donc nécessaire de se tourner vers des heuristiques pour obtenir des solutions approchées. Cependant il ne semble y avoir aucune heuristique garantie pour ce problème.

On dispose quand même d'un résultat intéressant, mais restreint au cas particulier d'une fonction de requête  $r$  constante, c'est-à-dire en fait quand l'évaluation se fait par l'indice de Shimmel et avec des relations de même poids entre paires de sommets. Dans ce cas, [Wu98]

7. la matrice origines/destinations est introduite section 1.1

8. On ne sait le résoudre qu'en examinant exhaustivement les (trop) nombreuses possibilités pour trouver une solution, ce qui dépasse très vite les capacités des ordinateurs quand la taille des données augmente. De plus les problèmes NP-complets sont liés : un algorithme efficace pour l'un le serait pour tous. Mais aucun algorithme n'a été trouvé en 30 ans pour un seul des milliers de ces problèmes. . .

donne un algorithme polynômial d'approximation. Plus précisément, l'algorithme calcule une  $(1 + \varepsilon)$ -approximation<sup>9</sup> en un temps  $\mathcal{O}\left(n^{2\lceil \frac{2}{\varepsilon} \rceil - 2}\right)$ . L'algorithme effectue une recherche exhaustive sur une famille d'arbres aussi adaptée que possible (choix non trivial!) : les *k-stars*, arbres ayant moins de  $k$  sommets de degré 2 ou plus. Plus le paramètre  $k$  est grand, plus le nombre d'arbres examinés, donc la qualité de l'approximation mais aussi le temps de calcul, le sont aussi.

Mais le problème *MCCST*, même sans se limiter à une fonction de requête constante, ne s'intéresse qu'à un ensemble de solutions réduit par rapport à ce qu'on cherche dans le cadre d'applications réelles. En effet le critère crucial pour un sous-réseau efficace n'est pas d'être un *arbre*, mais plutôt d'avoir un *coût faible* (longueur totale des arêtes) : la recherche doit porter sur les sous-graphes couvrant le graphe initial.

Dans cette voie, les travaux les plus intéressants semblent être ceux sur les *k-spanners*. Un sous-graphe couvrant  $\mathcal{G}'$  est un *k-spanner* d'un graphe  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  si :

$$\forall (u,v) \in \mathcal{V}^2, \quad \frac{d_{\mathcal{G}'}(u,v)}{d_{\mathcal{G}}(u,v)} \leq k$$

C'est à dire si la dilatation maximale (avec pour distance de référence la distance induite par le graphe initial  $\mathcal{G}$ ) est inférieure à  $k$ . Le problème du *sparsest k-spanner* est de trouver un *k-spanner* ( $k$  est donné) le plus économique (soit en nombre d'arêtes soit en longueur totale des arêtes). Ce problème est NP-complet, même pour des classes particulières de graphes (bipartis, planaires, triangulations . . .). [Kor99] montre aussi que même l'approximation en deçà d'un certain facteur d'erreur est un problème NP-complet. On peut cependant citer [ADDJ93], qui propose un algorithme polynômial pour trouver un  $2k + 1$ -spanner ayant moins de  $n \lceil n^{\frac{1}{k}} \rceil$  arêtes. De plus le poids total du sous-graphe ainsi trouvé est à un facteur  $\mathcal{O}\left(n^{\mathcal{O}(\frac{1}{k})}\right)$  de celui de l'arbre couvrant de poids minimum : on a une certaine garantie sur le coût. Cependant un *k-spanner* ne majore que la dilatation maximale, or si la dilatation moyenne est proche de  $k$  (qui est typiquement un entier supérieur à 2) le graphe obtenu est peu intéressant en pratique. Il faudrait donc préciser la qualité des résultats au regard de la dilatation moyenne. Idéalement, il faudrait aussi intégrer les pondérations de la matrice origines/destinations. Notons enfin que les recherches sur les *k-spanners* portent plutôt sur la minimisation du *coût* à efficacité majorée, alors que dans le cadre de construction ou protection selective dans les réseaux de transports, la condition d'efficacité semble prépondérante et il peut être préférable de chercher le réseau le plus efficace à coût fixé.

Ainsi la recherche de sous-réseaux, dans le cas général ou dans le cas où on se limite à des arbres, est difficile. Serait-il possible de trouver une classe de graphes, un peu dans l'idée des *k-stars* de [Wu98], ayant de bonnes propriétés de dilatation? On pense notamment aux triangulations (celle de Delaunay par exemple) vu leurs bonnes propriétés (voir section 2.3). Dans [Epp96], deux problèmes ouverts sont d'ailleurs posés à ce sujet :

- Une triangulation de dilatation minimale est-elle constructible en temps polynômial?
- Quelle est la pire dilatation de la triangulation de dilatation minimale?

En conclusion, de nombreuses recherches ont été menées sur des problèmes proches de ceux posés en 3.1, mais elles ne permettent pas encore d'obtenir des résultats satisfaisant en pratique, en particulier par des heuristiques adaptées et performantes.

---

9. c'est-à-dire que le ratio entre les indices de Shimmel de l'arbre obtenu et de l'arbre optimal (dont par ailleurs on ne sait rien) est inférieur à  $1 + \varepsilon$

### 3.3 Sous-réseaux renforcés

#### Problématique

La prévention des risques majeurs est une préoccupation importante des ministères de l'Environnement et de l'Équipement. Dans le contexte des réseaux de transport, les perturbations induites par des dommages potentiels ont conduit à mener un rapport d'évaluation sur la protection des réseaux. Un des objectifs de ce travail est permettre d'extraire des réseaux prioritaires, c'est-à-dire des sous-réseaux renforcés supposés être fonctionnels en toutes circonstances et garantir un service minimum en cas de crise. De tels réseaux pourraient être utilisés par les secours lors d'endommagements des infrastructures (ou par l'armée pour acheminer la paille en cas de canicule ayant fait fondre le macadam des routes). Le cas échéant, ces sous-réseaux seraient aussi reconstruits en priorité.

Ce projet correspond exactement au problème de détermination de sous-réseaux efficaces et économiques déjà évoqué. Le problème de décision associé (*Minimal Communication Cost Spanning Subgraph*) est formalisé page 20, et on montre qu'il est NP-complet. Le besoin de solutions, mêmes approchées, conduit donc à chercher des heuristiques.

Au passage, la détermination d'un tel réseau dépend directement de la mesure d'efficacité retenue. Elle est aussi liée, peut être plus subtilement, à la notion de vulnérabilité exposée section 1.3: les zones du réseau non retenues pour le réseau de crise sont *a priori* les moins vulnérables, *i.e.* celles dont la suppression ne diminue pas trop l'efficacité globale. Inversement, le réseau de crise concentre-t-il les zones vulnérables? Mieux cerner ces corrélations ainsi que la structure générale des points vulnérables pourrait conduire à des heuristiques plus rapides pour construire les sous-réseaux recherchés.

#### Un algorithme glouton

Les résultats qui seront présentés ci-dessous ont été obtenus par un algorithme de type glouton, *i.e.* qui fonctionne par étape, en optimisant simplement le passage d'une étape à la suivante mais sans stratégie d'optimisation à long terme. Ici le passage d'une étape à l'autre se fait en déterminant l'arête dont la suppression augmenterait le moins la dilatation, puis en supprimant effectivement cette arête. On donne page 21 un algorithme en  $\mathcal{O}(mn(n + m \ln n))$  pour déterminer cette arête<sup>10</sup>. En supposant de plus  $m = \mathcal{O}(n)$ , ce qui est le cas en pratique (le degré de ses carrefours ne croît pas avec la taille d'un réseau), on obtient finalement un algorithme en  $\mathcal{O}(n^4 \ln n)$ .

Bien évidemment les critères de sélection de l'arête à supprimer peuvent varier à l'infini: on peut supprimer l'arête minimisant n'importe quelle fonction du coût de cette arête et de l'augmentation de dilatation que sa suppression entraîne (mais il vaut mieux choisir raisonnablement la fonction en question!). Ici on veut privilégier la construction d'un réseau efficace, aussi le coût n'intervient pas dans le choix de l'arête à supprimer.

La non-optimalité de l'algorithme glouton ne saute pas aux yeux, mais on donne, page 21, un exemple de réseau où la solution retenue par cet algorithme n'est pas la meilleure. Ceci dit, la dilatation obtenue par le glouton dans cet exemple est à un facteur 1.22 de l'optimale, ce qui n'est pas ridicule pour un algorithme en  $\mathcal{O}(n^4 \ln n)$ <sup>11</sup>. Bien sûr ce n'est qu'un exemple: il faudrait savoir si l'erreur des solutions obtenues par le glouton est en général majorée. Si oui, quelle garantie peut-on obtenir sur l'heuristique? Si non, il faut construire un réseau générique pour lequel l'heuristique donne une solution arbitrairement loin de l'optimale. Ce travail reste à faire.

Il faut cependant reconnaître à un algorithme par étape un avantage pour la résolution du problème qui nous occupe. En effet, il y a deux contraintes sur le sous-réseau à obtenir: bonne

10.  $n$  (resp.  $m$ ) est le nombre de sommets (resp. d'arêtes) du graphe

11. à titre de comparaison, l'algorithme de [Wu98] donne en temps  $\mathcal{O}(n^4)$  une 1.66-approximation



efficacité et faible coût. En éliminant progressivement des arêtes, on augmente la dilatation et on diminue le coût, de sorte qu'un intervenant humain peut obtenir, à tâtons, un compromis dilatation/coût qui lui semble optimal.

### Résultats

On présente pour finir une application de l'algorithme glouton précédent sur une grille à mailles carrées et sur le réseau routier français.

Le réseau à mailles carrées met en évidence un phénomène de *hiérarchisation* : les arêtes sont ôtées de telle sorte que la structure initial revienne à plus grande échelle (mailles doubles en 11(c) par exemple). La figure 12(b) montre les variations de la dilatation au fur et à mesure qu'on enlève les arêtes.

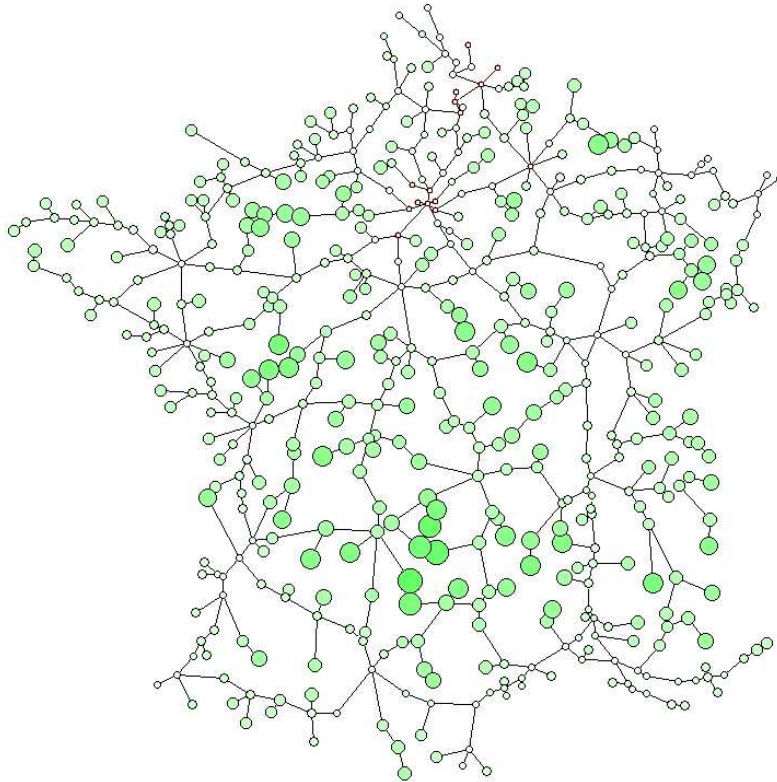


FIG. 10 – Sous-réseau constitué de 24% (en terme de longueur totale de routes) du réseau complet pour une augmentation de dilatation moyenne de 11%.

Cette hiérarchisation a aussi lieu sur le réseau routier français. Elle est naturellement facilitée par la hiérarchisation initiale du réseau (autoroutes ou routes). La figure 10 présente un sous-réseau intermédiaire entre le réseau initial et le dernier sous-réseau connexe (arbre couvrant). Pour une diminution de 76% de la longueur totale des routes, il limite l'augmentation de la dilatation moyenne à 11% (elle passe de 1.442 à 1.603). L'écart-type des dilatations relatives aux sommets augmente par contre plus (0.164 au lieu de 0.101 soit 62% d'augmentation), ce qui traduit un accroissement de inégalités sur le territoire. . . Continuer à élaguer ce réseau ne s'avère pas forcément judicieux : l'arbre obtenu juste avant de déconnecter le réseau diminue certes de 78% la longueur des routes du réseau initial, mais augmente de 48% la dilatation moyenne et de 350% l'écart-type : la dégradation du réseau n'est pas du tout linéaire en la diminution des coûts (voir figure 12(a)).

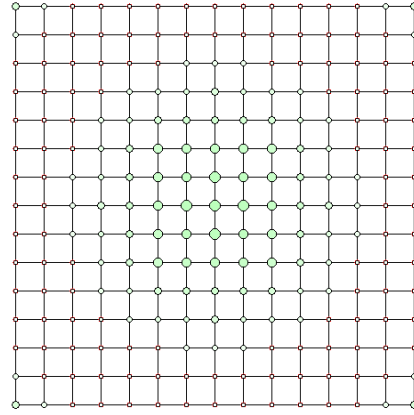
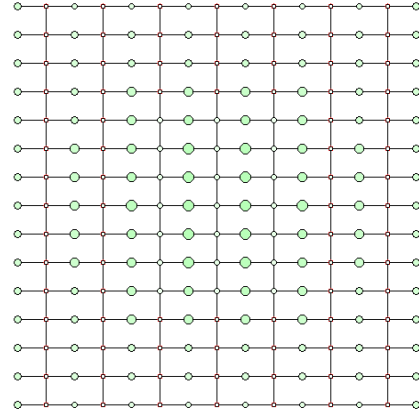
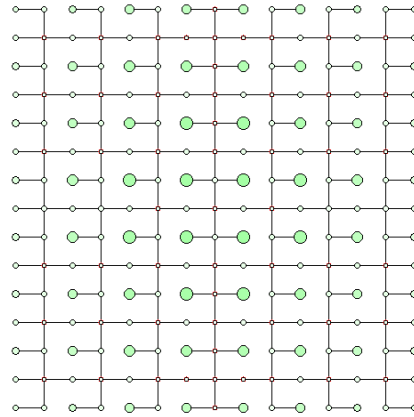
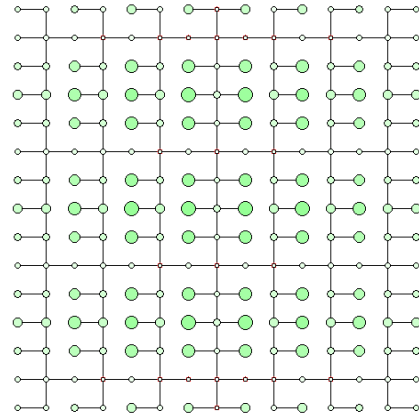
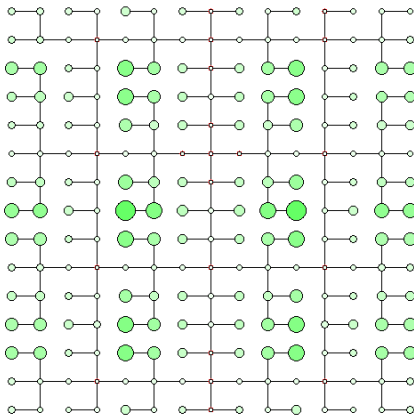
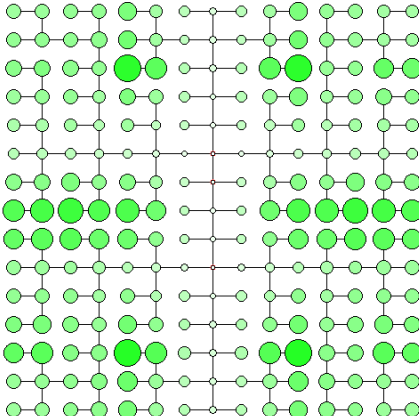
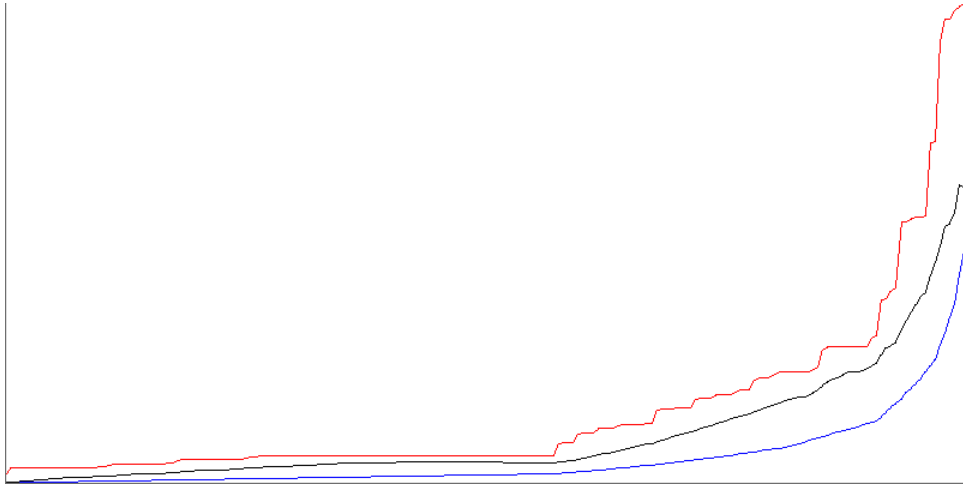
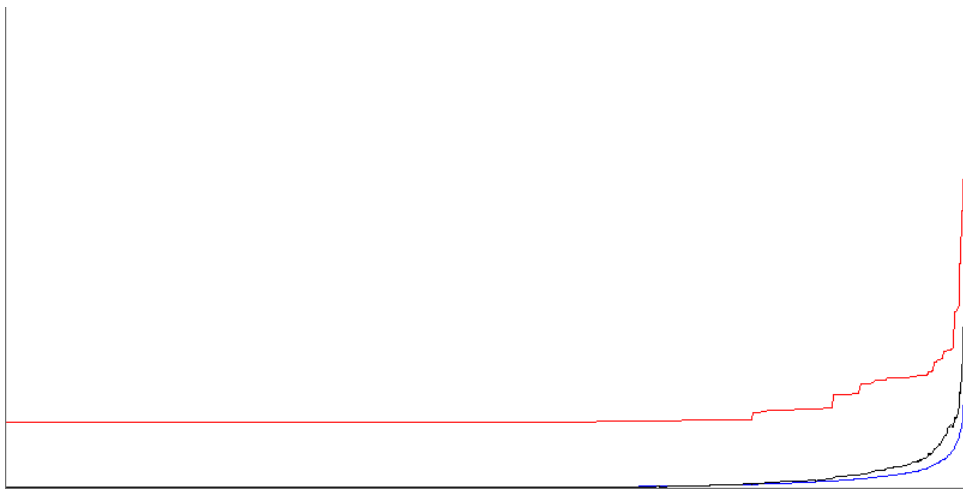
(a)  $m = 420$  ;  $\delta = 1.267$  ;  $\sigma = 0.008$ (b)  $m = 308$  ;  $\delta = 1.288$  ;  $\sigma = 0.023$ (c)  $m = 260$  ;  $\delta = 1.360$  ;  $\sigma = 0.068$ (d)  $m = 242$  ;  $\delta = 1.421$  ;  $\sigma = 0.095$ (e)  $m = 230$  ;  $\delta = 1.608$  ;  $\sigma = 0.170$ (f)  $m = 224$  ;  $\delta = 1.856$  ;  $\sigma = 0.219$ 

FIG. 11 – Étapes de sélection d'un sous-réseau d'une grille à mailles carrées. On donne le nombre d'arête  $m$ , la moyenne  $\delta$  et l'écart-type  $\sigma$  des dilatations des sommets.

(a) Lors de l'extraction sur une grille  $15 \times 15$  (figure 11)

(b) Lors de l'extraction sur le réseau routier français

FIG. 12 – Variations de la dilatation maximale (rouge), de la dilatation moyenne (bleu) et de l'écart-type (noir) en fonction de la longueur totale des arêtes enlevées par l'algorithme glouton.

## Conclusion

L'objectif initial du stage était très général : s'intéresser à l'évaluation des réseaux de transport, notamment dans un contexte de risques. Il m'a alors semblé naturel de chercher à quantifier la vulnérabilité des réseaux. Pourtant les recherches bibliographiques effectuées (dans le cadre de réseaux quelconques) ont conduit à des articles soit très généraux et loin d'être applicables, soit au contraire trop restreints : études de mesures particulières ayant un rapport plus ou moins direct, mais jamais vraiment explicite, avec les questions de vulnérabilité ou d'efficacité (indices de Kansky et Shimbel par exemple). J'ai donc été amené à tenter de définir ce qu'est un réseau vulnérable, et à cette fin à d'abord définir ce qu'est un réseau efficace. Cette partie, plus "géographique", prête sans doute le flanc aux critiques de spécialistes, mais c'est une bonne chose si cela peut conduire à combler le vide d'articles à ce sujet . . .

On a ensuite proposé une mesure d'efficacité particulière basée sur la dilatation. en s'attachant à montrer qu'elle semble correspondre plus que les mesures existantes à notre perception expérimentale de l'efficacité d'un réseau et qu'elle répond plutôt bien à la notion d'efficacité développée auparavant. Ce dernier point impose à une critique valable de préciser ou modifier les critères d'une mesure d'efficacité donnés en première partie et de montrer que la dilatation ne les satisfait plus.

On a alors pu envisager de trouver des réseaux efficaces pour cette mesure. On s'est en particulier intéressé au problème, répondant à une demande réelle, de déterminer un sous-réseau efficace d'un réseau donné. Malheureusement c'est un problème difficile (la NP-complétude du problème formalisé est démontrée) et les recherches actuelles en informatique ne semble pas offrir d'algorithmes d'approximation rapides et performants. On a proposé un algorithme qui a l'avantage de fournir des solutions, qui de plus semblent assez bonnes. Mais on n'a par contre pas pu déterminer des bornes sur la qualité des approximations dans les pires cas.

Le cadre peu exploré et assez large du stage a permis de soulever beaucoup de questions, de s'intéresser à certaines sans pouvoir toujours aller au bout et d'en négliger d'autres. Parmi les points importants dont il faudrait poursuivre l'étude, il y a bien sûr le problème de définition et formalisation des concepts d'efficacité et vulnérabilité.

On peut aussi recentrer plus l'étude sur la vulnérabilité des réseaux : l'efficacité occupe une grande partie du rapport, même si on a cherché à définir la vulnérabilité et si la protection de sous-réseaux efficaces est aussi liée à la vulnérabilité. Autour des courbes de vulnérabilité (section 1.3), on peut chercher à obtenir celles de réseaux (réels ou types) ou étudier d'éventuels phénomènes de seuil (brusque augmentation de rapidité au-delà d'un certain niveau d'endommagement). Un autre phénomène juste entrevu et non exposé ici est celui des "reports de vulnérabilité" : comment l'endommagement d'un réseau en un point vulnérable redistribue les déplacements et définit ainsi de nouveaux points vulnérables ?

D'un point de vue plus informatique, comment obtenir des algorithmes d'approximation performants pour le problème de la détermination d'un sous-réseau efficace ? Peut être par une démarche similaire à celle de [Wu98] avec les  $k$ -stars (section 3.2) : trouver une classe restreinte de sous-réseaux approximant assez bien (ceci étant à quantifier) l'optimum et choisir exhaustivement parmi ces sous-réseaux le meilleur.

Mais la difficulté de ce stage et de ses éventuelles suites tient peut être surtout à ce qu'il s'attache à résoudre des problèmes pratiques dont les objectifs ne sont cependant pas toujours clairement définis. . .

## A Annexe

### A.1 Densité angulaire d'une surface

On définit la *densité angulaire*  $p_S$  d'une partie  $S$  de  $\mathbb{R}^2$  par la densité de probabilité de la pente d'un segment défini par deux points tirés au hasard sur  $S$ .

Plaçons nous dans le cas de surfaces raisonnables, notamment à contours dérivables sauf peut-être en un nombre fini de points. Avec les notations de la figure 13 : si on fixe un point  $(x,y)$  sur une surface  $S$ , la probabilité qu'un deuxième point tiré au hasard sur  $S$  définisse une pente entre  $\alpha$  et  $\alpha + d\alpha$  est proportionnelle à l'aire de la surface grisée. Cette dernière vaut (à l'ordre 1 en  $d\alpha$ , pour peu qu'on ne tombe pas sur un coin du contour) :  $\pi(p(x,y)^2 + q(x,y)^2) \frac{d\alpha}{2\pi}$ . En sommant pour  $(x,y) \in S$ , on obtient un terme de la forme  $f(\alpha)d\alpha$ . La densité angulaire  $p_S$  est alors proportionnelle à  $f$ .

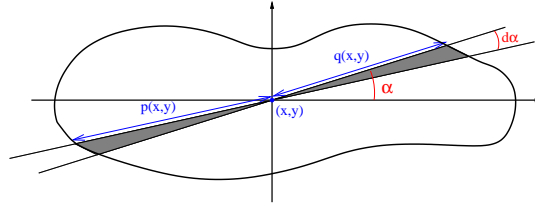


FIG. 13 – Calcul de la densité angulaire d'une surface  $S$

- pour un disque, on obtient une densité angulaire constante (pas d'angle favorisé).
- pour un triangle équilatéral :

$$\forall \alpha \in \left[0, \frac{\pi}{3}\right], \quad p_S(\alpha) = \lambda \frac{1}{(\sqrt{3} \cos(\alpha) + \sin(\alpha))^3}$$

On étend  $p_S$  à  $[0, \pi]$  par  $\frac{\pi}{3}$ -périodicité (Figure 14).

- pour un carré :

$$\forall \alpha \in \left[0, \frac{\pi}{4}\right], \quad p_S(\alpha) = \lambda \frac{2 - \tan(\alpha)}{\cos(\alpha)^2}$$

On étend  $p_S$  à  $[0, \pi]$  par symétrie d'axe  $\alpha = \frac{\pi}{4}$  et par  $\frac{\pi}{2}$ -périodicité (Figure 15).

- pour une surface quelconque, on se contente en général d'approcher le résultat par un processus de Monte-Carlo.

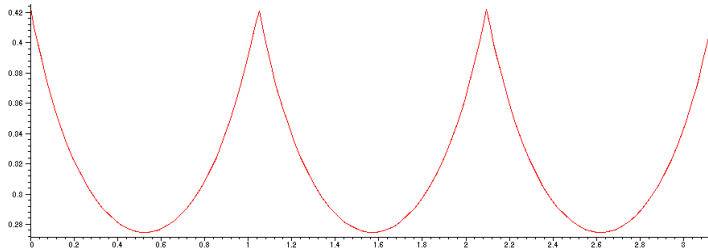


FIG. 14 – Densité angulaire d'un triangle équilatéral (un des côtés définit la pente nulle).  
Maxima en  $k\frac{\pi}{3}$ , minima en  $\frac{\pi}{6} + k\frac{\pi}{3}$ .

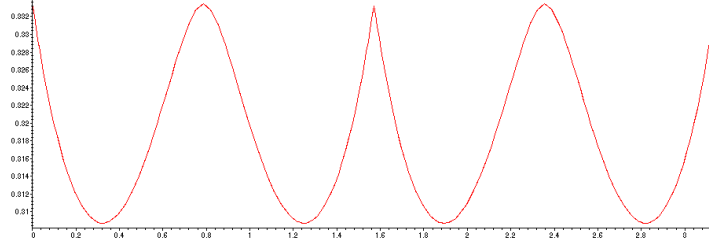


FIG. 15 – Densité angulaire d'un carré (un des côtés définit la pente nulle).  
Maxima en  $k\frac{\pi}{4}$ , premier minimum en  $\arctan\frac{1}{3}$ .

## A.2 Un résultat de NP-complétude

On appelle *Minimal Communication Cost Spanning Subgraph (MCCSS)* le problème suivant :

Étant donnés deux réels strictement positifs  $c$  et  $\delta$ , un graphe  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  pondéré par  $w : \mathcal{E} \mapsto \mathbb{R}_*^+$  et une fonction de requête  $r : \mathcal{V}^2 \mapsto \mathbb{R}_*^+$ , existe-t-il un sous-graphe  $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$  de  $\mathcal{G}$  tel que :

$$\sum_{e \in \mathcal{E}'} w(e) \leq c \quad (\text{Condition de coût})$$

$$\sum_{u,v \in \mathcal{V}}^{u \neq v} W(u,v)r(u,v) \leq \delta \quad (\text{Condition d'efficacité})$$

Où  $W(u,v)$  est le minimum des poids des chemins de  $\mathcal{G}'$  reliant  $u$  et  $v$ .

Montrons par réduction à *MCCST* la NP-complétude de ce problème.

Vérifier qu'un sous-graphe donné est solution de *MCCSS* se fait clairement en temps polynomial : *MCCSS* est dans NP.

Soit alors une instance  $(\delta, \mathcal{G} = (\mathcal{V}, \mathcal{E}), w, r)$  de *MCCST*. On pose :

$$\begin{aligned} K &= \sum_{e \in \mathcal{E}} w(e) \\ \tilde{w}(e) &= w(e) + K \\ c &= K|\mathcal{V}| \\ \tilde{\delta} &= \delta + K \left( \sum_{u,v \in \mathcal{V}}^{u \neq v} r(u,v) \right) \end{aligned}$$

Montrons qu'une solution  $\mathcal{G}'$  du problème *MCCSS* d'instance  $(c, \tilde{\delta}, \mathcal{G}, \tilde{w}, r)$  est solution de *MCCST*.

$\mathcal{G}'$  est connexe sinon pour deux sommets  $u$  et  $v$  non reliés,  $\tilde{W}(u,v) = \infty$  ce qu'exclut la satisfaction de la condition d'efficacité de *MCCSS*. De plus,  $\mathcal{G}'$  ne peut avoir strictement plus de  $|\mathcal{V}| - 1$  arêtes, car comme  $\tilde{w}(e) > K$ , on aurait un poids total des arêtes strictement supérieur à  $c = K|\mathcal{V}|$ , ce qu'exclut la condition de coût.  $\mathcal{G}'$  est donc un arbre.

Par ailleurs, pour  $u \neq v$  on a  $\tilde{W}(u,v) \geq W(u,v) + K$ , d'où :

$$\tilde{\delta} \geq \sum \tilde{W}(u,v)r(u,v) \geq \sum W(u,v)r(u,v) + K \sum r(u,v)$$

Comme  $\tilde{\delta} = \delta + K \sum r(u,v)$ , on a  $\sum W(u,v)r(u,v) \leq \delta$  et  $\mathcal{G}'$  est donc solution de *MCCST*.

De plus l'instance de *MCCSS* ainsi définie se calcule en temps polynômial et est de taille polynômiale en la taille de l'instance de départ. *MCCST* n'est donc pas plus difficile que *MCCSS* : trouver une solution pour *MCCSS* en temps polynômial permet de le faire pour *MCCST*. Comme *MCCST* est NP-complet, *MCCSS* l'est aussi.

### A.3 Exemple de non-optimalité de l'algorithme glouton

On expose ici un exemple de réseau pour lequel la stratégie gloutonne (consistant à enlever à chaque étape l'arête dont la suppression augmente le moins la dilatation) ne conduit pas à la meilleure solution pour le problème du sous-graphe efficace.

Considérons le réseau constitué d'un triangle équilatéral de côté  $\sqrt{3}$  et de son centre de gravité relié aux sommets du triangle (par des segments de longueur 1 donc). Ce réseau est celui numéroté 1 figure 16. Cette même figure présente les différentes stratégies pour élaguer le réseau (on représente par un seul réseau toutes ses variantes obtenues par symétrie ou rotation).

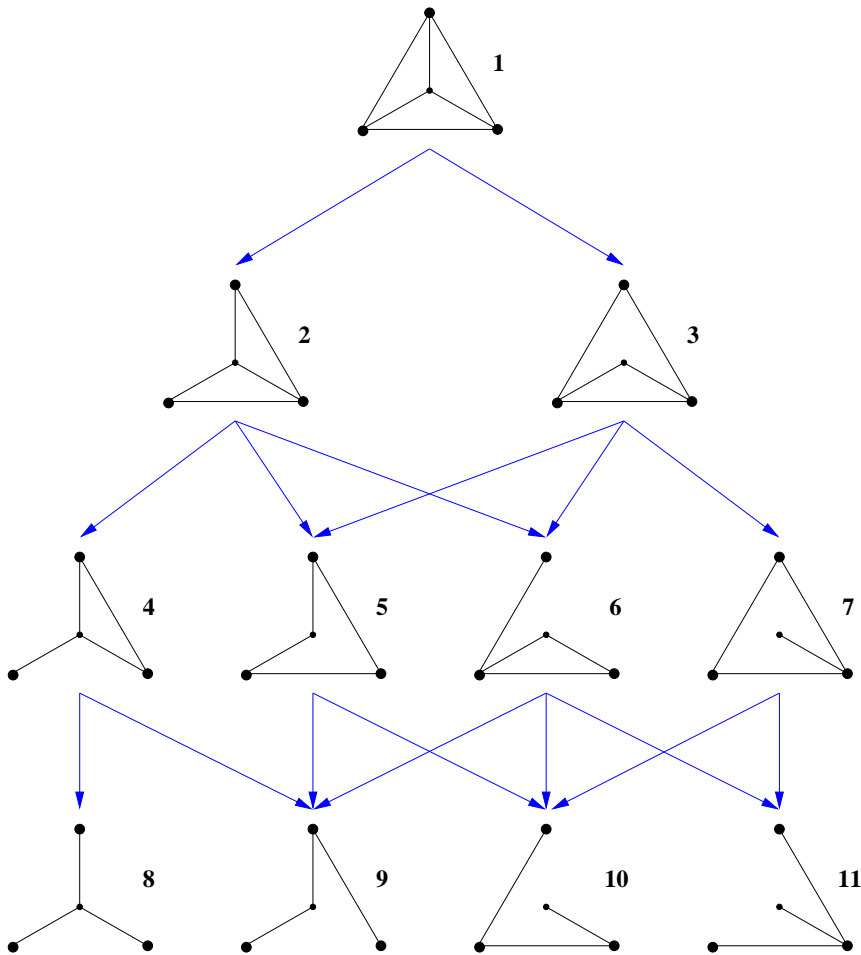


FIG. 16 – Différentes stratégies de sélection d'un sous-réseau

On suppose de plus que les relations entre les sommets du triangle ont un poids  $q$ , tandis que les liaisons entre le centre de gravité et ces mêmes sommets n'ont qu'un poids 1 (si on ne veut se passer de la matrice origines/destinations, on peut imaginer que les trois sommets du triangle sont eux-mêmes des réseaux de plusieurs sommets).

Calculons alors la dilatation moyenne (en fonction de  $q$  et pour  $q = 12$ ) et le coût (en terme de longueur totale d'arêtes) des réseaux de la figure 16 :

réseau	$(3q + 3)\delta(q)$	$\delta(12)$	coût
1	$3q + 3$	1.0000	$3 + 3\sqrt{3} \simeq 8.2$
2	$2q(1 + \frac{1}{\sqrt{3}}) + 3$	1.0476	$3 + 2\sqrt{3} \simeq 6.5$
3	$3q + 3 + \sqrt{3}$	1.0444	$2 + 3\sqrt{3} \simeq 7.2$
4	$q(1 + \frac{4}{\sqrt{3}}) + 3$	1.0952	$3 + \sqrt{3} \simeq 4.7$
5	$2q(1 + \frac{1}{\sqrt{3}}) + 3 + \sqrt{3}$	1.0920	$2 + 2\sqrt{3} \simeq 5.5$
6	$4q + 3 + \sqrt{3}$	1.3521	$2 + 2\sqrt{3} \simeq 5.5$
7	$3q + 3 + 2\sqrt{3}$	1.0888	$1 + 3\sqrt{3} \simeq 6.2$
8	$2\sqrt{3}q + 3$	1.1428	3
9	$2q(1 + \frac{2}{\sqrt{3}}) + 3 + \sqrt{3}$	1.4473	$2 + \sqrt{3} \simeq 3.7$
10	$4q + 3 + 3\sqrt{3}$	1.4409	$1 + 2\sqrt{3} \simeq 4.5$
11	$4q + 3 + 2\sqrt{3}$	1.3965	$1 + 2\sqrt{3} \simeq 4.5$

La valeur  $q = 12$  a été choisie à dessein : c'est en effet au delà de cette valeur qu'on observe une plus petite dilatation pour le réseau 3 que pour le 2. Si on exécute alors l'algorithme glouton (avec un coût maximal de 5 par exemple) les réseaux successivement obtenus seront les réseaux 1, 3, 7 et 11. Pourtant passer par les réseaux 1, 2, 4 et 8 (donc une stratégie alternative) conduit à une solution meilleure (le réseau 8 a une dilatation plus faible que le 11). Ceci prouve la non-optimalité de l'algorithme glouton.

#### A.4 Minimiser l'augmentation de dilatation

Soit un graphe  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  pondéré par  $w : \mathcal{E} \mapsto \mathbb{R}_*^+$ . Pour  $\mathcal{E}' \subset \mathcal{E}$ , on note  $d_{\mathcal{E}'}$  la distance sur le graphe  $(\mathcal{V}, \mathcal{E}')$  et  $\delta_{\mathcal{E}'}$  la dilatation moyenne. On veut déterminer l'arête  $e^* \in \mathcal{E}$  dont la suppression augmente le moins la dilatation moyenne, *i.e.* telle que

$$\delta_{\mathcal{E} - \{e^*\}} = \min_{e \in \mathcal{E}} \delta_{\mathcal{E} - \{e\}}$$

On pose  $n = |\mathcal{V}|$  et  $m = |\mathcal{E}|$ .

**Un algorithme en  $\mathcal{O}(mn(n + m \ln n))$**

L'algorithme de Dijkstra implémenté par tas de Fibonacci (voir [CLR92]) calcule en  $\mathcal{O}(n + m \ln n)$  les distances d'un sommet donné aux autres sommets. En répétant ce calcul pour tous les sommets, on obtient les plus courtes distances entre paires de sommets du graphe en temps  $\mathcal{O}(n(n + m \ln n))$ , puis on en déduit la dilatation en temps  $(n^2)$ , ce qui n'augmente pas la complexité.

On détermine alors simplement l'arête  $e^*$  minimisant l'augmentation de dilatation : il suffit, pour chacune des  $m$  arêtes, de supprimer la supprimer du graphe, de calculer la dilatation, de restaurer l'arête et de passer à la suivante. L'algorithme est clairement en  $\mathcal{O}(mn(n + m \ln n))$ .

#### Le même algorithme un peu amélioré

Le gros défaut de l'algorithme précédent est de calculer un nombre effarant de fois la même chose. D'une part seule une partie des sommets voient leur distance affectée par la rupture d'un arête (il faut qu'ils n'aient qu'un unique plus court chemin, passant par cette arête). Donc il est superflu de relance Dijkstra sur *tous* les sommets après la rupture d'une arête : on peut déterminer en  $\mathcal{O}(n)$  les sommets auxquels on peut se restreindre. D'autre part, lorsqu'une arête est supprimée, les plus courts chemins qui passaient par cette arête vont être



redistribués sur le graphe. Mais intuitivement, et la pratique le confirme, seule une partie des arêtes restantes va être traversée par ces nouveaux plus courts chemins. Aussi lorsque on aura déterminé l'arête  $e^*$ , on détermine les arêtes qui récupèrent des plus courts chemins lors de la rupture de  $e^*$  (ce qui se fait en  $\mathcal{O}(mn^2)$ , *i.e.* sans augmenter la complexité) : lors de l'étape suivante on pourra se limiter à recalculer l'augmentation de dilatation pour la rupture de ces seules arêtes (pour les autres arêtes, avoir cassé  $e^*$  n'a pas changé l'augmentation de dilatation que leur propre rupture pouvait induire).

Ces améliorations ne changent pas la complexité asymptotique, ou du moins nous n'avons pas trouvé de bornes intéressantes, pour des réseaux quelconques, du nombre de sommets et d'arêtes concernés.

### Un algorithme en $\mathcal{O}(mn^3)$

On présente ici un algorithme qui, bien que la majoration obtenue sur sa complexité soit moins bonne que pour l'algorithme précédent, s'exécute en pratique beaucoup plus rapidement.

La constatation qui l'a inspiré est la suivante: quand une arête casse, les nouveaux plus courts chemins sont souvent redistribués de la même manière sur le graphe, c'est-à-dire que seules quelques arêtes, jouant en quelque sorte le rôle de celle supprimée, vont canaliser ces plus courts chemins. Nous allons formaliser ceci.

Pour une arête  $AB$ , on note  $\mathcal{S}_A$  l'ensemble des sommets tels que tout plus court chemin entre un de ces sommets et le sommet  $B$  passe par  $A$ . L'ensemble  $\mathcal{S}_B$  est défini symétriquement. Si un plus court chemin de  $P$  à  $Q$  passe par l'arête  $AB$  (figure 17(a)), sa restriction de  $P$  à  $B$  est un plus court chemin : si un chemin plus court reliait  $P$  à  $B$  (figure 17(b)), il se prolongerait en un chemin de  $P$  à  $Q$  plus court (figure 17(c)). C'est donc que  $P$  est dans  $\mathcal{S}_A$ . Symétriquement,  $Q$  est dans  $\mathcal{S}_B$ . Ainsi les couples de  $\mathcal{S}_A \times \mathcal{S}_B$  sont ceux affectés (*i.e.* la distance entre les deux sommets change) par la rupture de l'arête  $AB$ .

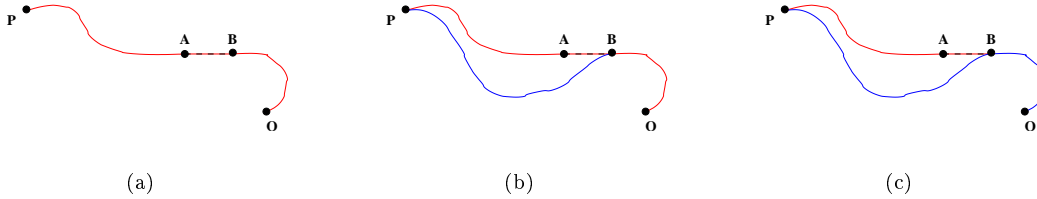


FIG. 17 – Caractérisation des ensembles  $\mathcal{S}_A$  et  $\mathcal{S}_B$

On appelle *séparateur* de  $\mathcal{S}_A$  et  $\mathcal{S}_B$  un ensemble  $\mathcal{S}$  de sommets<sup>12</sup> disjoint de  $\mathcal{S}_A$  et  $\mathcal{S}_B$  et tel que tout plus court chemin entre  $\mathcal{S}_A$  et  $\mathcal{S}_B$  passe par un de ces sommets.

On a alors la propriété intéressante suivante: si la distance entre deux sommets  $P$  et  $Q$  est modifiée par la rupture de l'arête  $AB$ , alors  $P \in \mathcal{S}_A$  et  $Q \in \mathcal{S}_B$  (ou l'inverse), et la nouvelle distance entre ces sommets est la longueur d'un plus court chemin passant par un sommet  $R$  de  $\mathcal{S}$ . De plus, comme  $R \notin \mathcal{S}_A \cup \mathcal{S}_B$ , les distances  $d_{\mathcal{E}}(P, R)$  et  $d_{\mathcal{E}}(R, Q)$  ne sont pas affectées par la rupture de  $AB$ . On a donc en fait :

$$d_{\mathcal{E}-\{AB\}}(P, Q) = \min_{R \in \mathcal{S}} (d_{\mathcal{E}}(P, R) + d_{\mathcal{E}}(R, Q))$$

12. En vérité, on est parfois obligé de prendre un "bi-sommet" (deux sommets reliés par une arête du graphe) à cheval sur  $\mathcal{S}_A$  et  $\mathcal{S}_B$  lorsque ces deux ensembles sont adjacents (il y a des arêtes directes entre eux). Sinon quelques cas pathologiques empêchent de réutiliser les plus courtes distances déjà connues. Si vous n'avez rien compris c'est normal: il faudrait une bonne page pour expliquer ce détail obscur et sans grand intérêt.

L'algorithme procède alors ainsi pour calculer l'augmentation de dilatation en cas de rupture de l'arête  $AB$  :

1. il détermine  $\mathcal{S}_A$  et  $\mathcal{S}_B$  en temps  $\mathcal{O}(n)$  en remarquant :  $P \in \mathcal{S}_A \Leftrightarrow d(P,B) = d(P,A) + d(A,B)$  ;
2. il choisit  $\mathcal{S}$  : l'adjacence de  $\mathcal{S}_A$  par exemple se calcule en temps  $\mathcal{O}(mn)$  ;
3. Recalcule la distance au sein des couples de  $\mathcal{S}_A \times \mathcal{S}_B$  avec la formule du minimum sur  $\mathcal{S}$  précédente en temps  $|\mathcal{S}_A||\mathcal{S}_B||\mathcal{S}| = \mathcal{O}(n^3)$ .

Comme il y a  $\mathcal{O}(m)$  arêtes à examiner ainsi, on a bien la complexité annoncée.

### Comparaison pratique des algorithmes

On appellera dans la suite algorithme 1 celui en  $\mathcal{O}(mn(n + m \ln n))$  (avec amélioration sur le nombre d'arêtes considérées) et algorithme 2 celui en  $\mathcal{O}(mn^3)$ . La figure 18 représente l'évolution du temps de calcul à chaque étape du glouton.

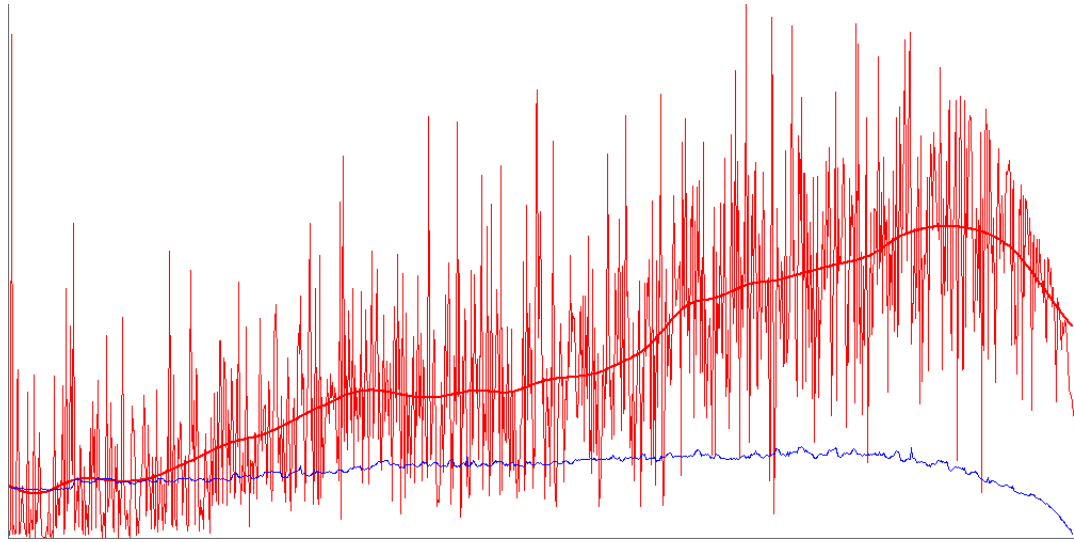
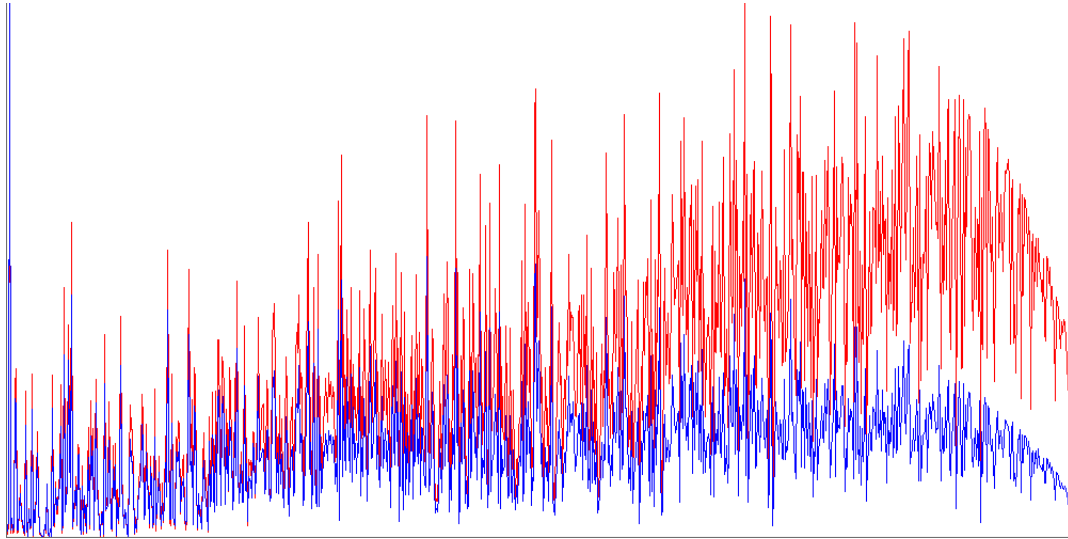


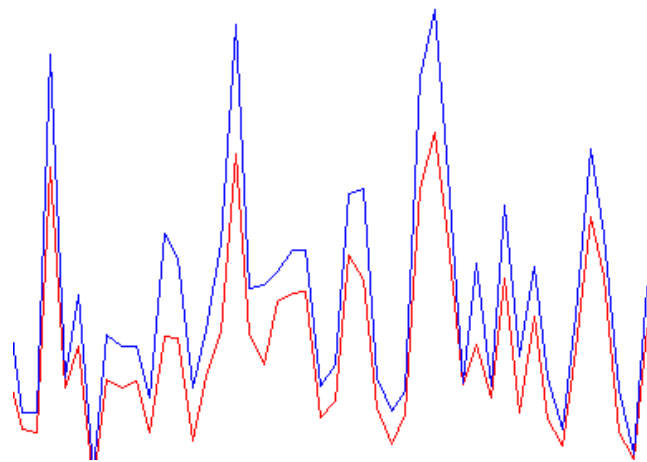
FIG. 18 – *Comparaison des temps de calcul à chaque étape du glouton (en rouge l'algorithme 1, en bleu le 2)*

La figure 19 montre, dans le cas de l'algorithme 1, la corrélation (normale!) entre le temps de calcul d'une étape et le nombre restreint d'arêtes que l'algorithme considère à cette étape.

La figure 20 montre une corrélation plus surprenante entre le temps de calcul d'une étape et le plus petit cardinal des ensembles  $\mathcal{S}_A$  et  $\mathcal{S}_B$ . Je n'ai pas eu le temps de chercher à comprendre d'où venait cette corrélation, alors que vu la complexité de l'algorithme une corrélation du temps de calcul avec le produit  $|\mathcal{S}_A||\mathcal{S}_B||\mathcal{S}|$  semble plus logique (mais elle n'existe pas). Ceci dit, comme cet algorithme est en pratique plus rapide que l'algorithme 1 malgré une borne de complexité moindre, peut-être est-ce un espoir d'améliorer cette borne. . .

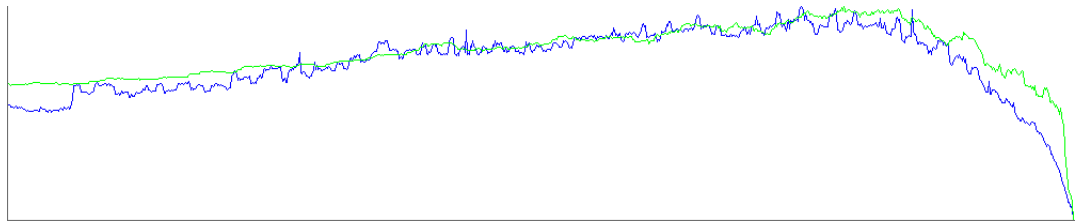


(a) corrélation globale bonne (0.85)

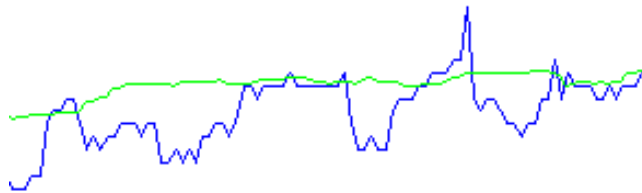


(b) corrélation locale très forte (0.98 sur 100 arêtes)

FIG. 19 – *algorithme 1* : corrélation entre temps de calcul (rouge) et nombre d'arêtes considérées (bleu)



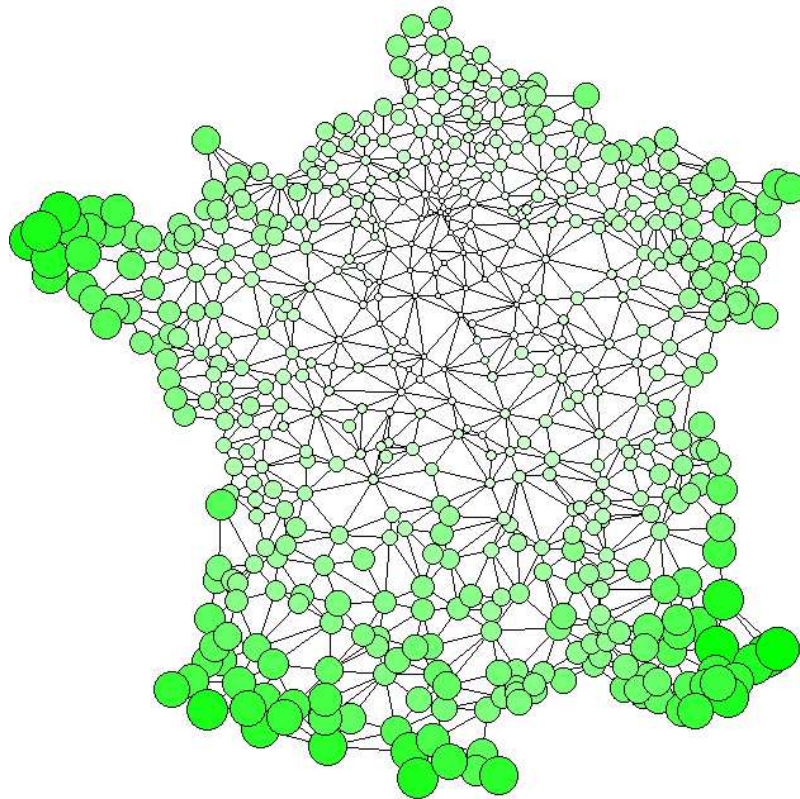
(a) corrélation globale forte (0.93)



(b) corrélation locale moyenne (0.63 sur 100 arêtes)

FIG. 20 – *algorithme 2 : corrélation entre temps de calcul (bleu) et  $\min(|S_A|, |S_A|)$  (vert)*

## A.5 Indices de Shimbel et Kansky sur le réseau routier français

FIG. 21 – *Indice de Shimbel.*

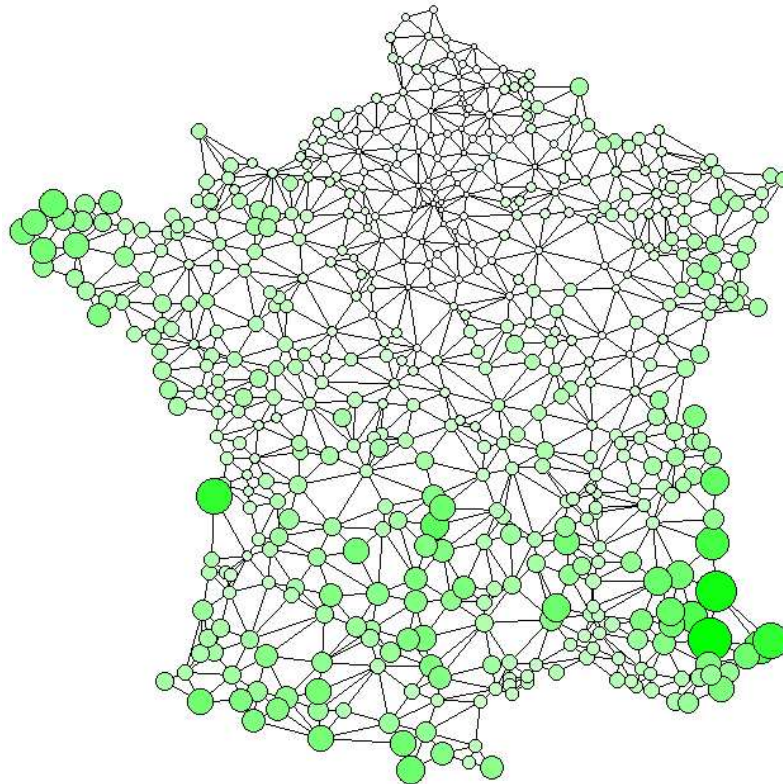


FIG. 22 – *Indice de Kinsky.*

## Références

- [ADDJ93] I. Althöfer, G. Das, D. Dobkin, D. Joseph *Generating sparse spanners for weighted graphs*. Discrete Computational Geometry, 9, 1993, pp 89-100.
- [BR97] S. Bulteau et G. Rubino. *Evaluating network vulnerability with the mincuts frequency vector*. Publication interne 1089, IRISA, 1997.
- [BTZ98] F. Baccelli, K. Tchoumatchenko, S. Zuyev *Markov-paths on the Poisson-Delaunay graph*. Publication interne 3420, INRIA, 1998.
- [BY95] J. D. Boissonnat, M. Yvinec *Géométrie algorithmique*. Ediscience international, Paris, 1995.
- [Che89] L. P. Chew *There is a planar graph almost as good as the complete graph*. Proc. 2nd ACM Symp.Comp.Geom., 1986, pp 169-177.
- [CLR92] T. H. Cormen, C. E. Leiserson, R. L. Rivest *Introduction to algorithms*. MIT Press, 1992.
- [Epp96] D. Eppstein *Spanning trees and spanners*. Tech. Report 96-16 (University of California, Irvine), 1996.
- [Gle03] J. F. Gleyze *Fonctionnalité du métro parisien - Efficacité et fiabilité du réseau de métro parisien dans l'éventualité de catastrophes*. Actes des rencontres Theo-Quant, Besançon, 2003.
- [Kan89] K. Kansky *Measures of network structure*. Flux, 1989
- [KG89] J. M. Keil et C. A. Gutwin *The Delaunay triangulation closely approximates the complete Euclidean graph* Proc. 1st Worksh Algorithms and Data Structures. Springer LNC 382, 1989, pp 47-56.
- [Kor99] G. Kortsarz *On the hardness of approximating spanners*. 1999.
- [Shi53] A. Shimbel *Structural parameters of communication networks*. Bulletin of Mathematical Biophysics 15, 1953, pp 501-507.
- [Wu98] B. Y. Wu, G. Lancia, V. Bafna, K. Chao, R. Ravi, C. Y. Tang *A Polynomial Time Approximation Scheme for Minimum Routing Cost Spanning Trees*. 1998.