

TIPE : Etude de couvertures de réseaux de métros, application de l'homologie persistante

Elowan
elowarp@gmail.com

Contents

I Définitions	1
I.1 Constructions géométriques	2
II Méthode	4
III Les données	6
III.1 Sources	6
III.2 Construction des informations importantes	6
IV Résultats et conclusion	7
Bibliography	8

Abstract

Nous nous proposons ici d'étudier les différentes disparités dans les réseaux métropolitains de plusieurs grandes villes, nous voulons détecter les zones spatiales les plus en déficit de transports en commun. Pour cela, nous adopterons une approche analytique utilisant la topologie, *l'homologie persistante*.

I Définitions

Note: Est ce que c'est considéré comme du plagiat ? (Extrait notice XENS : Les textes et figures sont l'œuvre du candidat. Les reproductions et les copies ne sont pas acceptées, tout plagiat (« action d'emprunter un passage de tout auteur en les donnant pour sien. ») est une forme de contrefaçon et constitue un délit et donc une fraude, il est susceptible d'être sanctionné en tant que telle suivant la procédure disciplinaire.)

L'exemple suivant est tiré de [1].

Prenons comme exemple le style artistique du pointillisme. Lorsque l'on regarde une oeuvre d'art, comme le tableau de Seurat en Figure 1, nous voyons bien plus qu'un grand nombre de points, nous voyons des formes et des objets. D'une discrétisation, nous en faisons un *continuum* de formes.

L'homologie persistante va, à partir d'une discrétisation, chercher à fournir un suivi d'évolution des formes que l'on pourrait distinguer de cet ensemble.

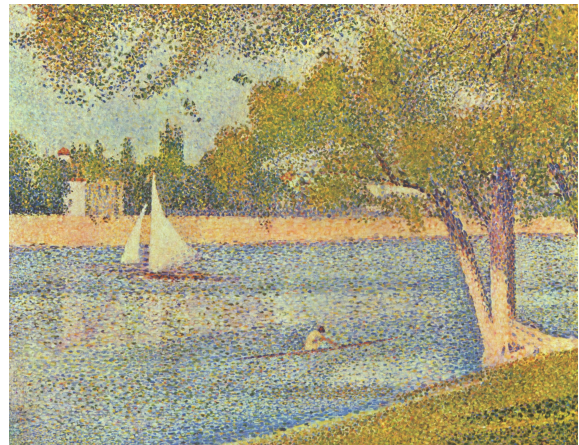


Figure 1: La scène à la Grande Jatte - Printemps (Georges Seurat, 1888)

Nous chercherons ici seulement à caractériser les “trous” dans un espace, afin de détecter les trous de couverture dans un réseau métropolitain, étant défini ici comme un espace de \mathbb{R}^2 où chaque point correspond à une station de métro.

I.1 Constructions géométriques

Afin d'utiliser l'homologie persistante, nous devons définir certaines notions géométriques, nous noterons dans la suite X l'ensemble des points de \mathbb{R}^2 que l'on considère.

Définition :

Un simplexe σ de dimension k correspond à l'enveloppe convexe de $k + 1$ points de X non inclus dans un même sous-espace affine de dimension $k - 1$.

On définit un simplexe de dimension 0 comme un point de X .

Par exemple, un simplexe de dimension 1 est un segment et un simplexe de dimension 3 est un trièdre.

Remarque : On dit que σ_i est une face de σ_j si et seulement si $\sigma_j \subset \sigma_i$ et la dimension de σ_i est strictement supérieure à celle de σ_j .

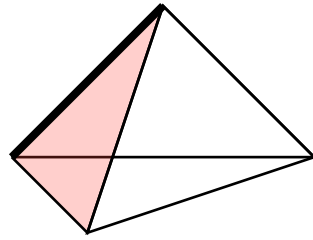


Figure 2: Ce trièdre est un simplexe σ de dimension 3, où le triangle rouge représente un simplexe τ de dimension 2 mais aussi une face de σ dans le sens de la remarque précédente.

Notons que l'arête en gras est un simplexe de dimension 1 et est une face de σ et τ .

Définition :

Un *complexe simplicial* est un ensemble de simplexes.

Définition :

Une *filtration* est une application qui à un entier i associe un complexe simplicial K_i de telle sorte que $\forall j \in \llbracket 0, i \rrbracket, K_j \subset K_i$

Observons sur Figure 3 les trois notions précédemment définies. Chaque K_i est un complexe simplicial, la suite $(K_i)_{i=0}^4$ est une filtration et tous les points, segments et faces (ici grisées) sont des simplexes de dimension 0, 1 et 2 respectivement.

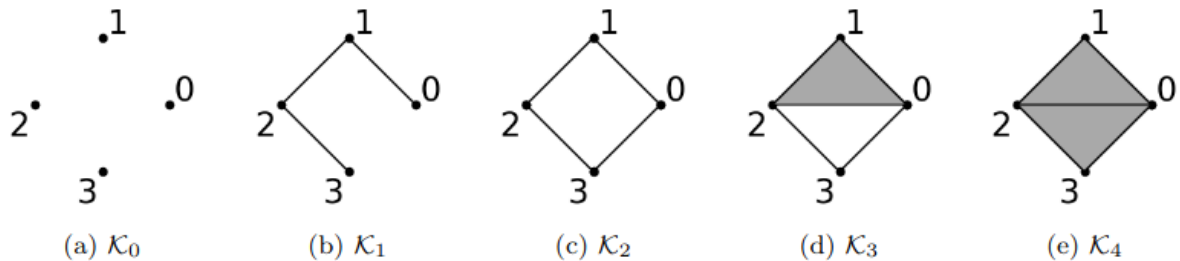


Figure 3: Représentation d'une filtration où $K_0 \subset K_1 \subset K_2 \subset K_3 \subset K_4$, tiré de [2] (à refaire)

Nous observons qu'une filtration permet d'ajouter une notion de "temporalité" dans un ensemble de points. Nous sommes capable de noter quels événements surviennent entre deux

complexes à la suite, par exemple l'apparition d'un cycle entre K_1 et K_2 (le cycle (0,1,2,3)) ou l'apparition d'un simplexe de dimension 2 entre K_2 et K_3 (la face grisée).

On introduit de plus un ordre total $<$ sur l'ensemble des simplexes d'une filtration.

Définition :

Soient une filtration $K_0 \subset K_1 \subset \dots \subset K_p$ et deux simplexes $\sigma \in K_i$ et $\tau \in K_j$, on a

$$\begin{cases} \text{soit } i < j \\ \text{sinon } i = j \text{ et } \sigma \text{ est une face de } \tau \end{cases} \Rightarrow \sigma < \tau$$

Si aucun des deux cas n'est réalisé alors le choix de l'ordre entre les deux simplexes est arbitraire.

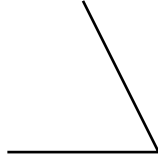


Figure 4: Exemple de filtration avec un ordre défini sur celui ci, pas terminé

Il y a un problème : nous voulons analyser un ensemble de points discrets, et non pas une filtration déjà existante, il nous faut alors créer une filtration depuis un ensemble de points. Cela va se faire via une construction incrémentale de complexes simpliciaux via les complexes de Vietoris-Rips pondérés. Ainsi d'après [2] :

Définition :

Soient un ensemble $X = (x_i)_{i=0}^n$ de points associés à des poids $(w_i)_{i=0}^n$ et une distance d , on définit le complexe simplicial pondéré de Vietoris-Rips au rang r , noté $V_r(X, d)$, comme l'ensemble des simplexes $(x_{i_0}, \dots, x_{i_k})$ tels que :

$$\begin{cases} \forall j \in \llbracket 0, k \rrbracket, w_{i_j} < r \\ \forall (j, l) \in \llbracket 0, k \rrbracket^2, d(x_{i_j}, x_{i_l}) + w_{i_j} + w_{i_l} < 2r \end{cases}$$

Ainsi plus on augmente r , plus le complexe possède des simplexes, on en donne une représentation Figure 5. (*Note : Remarque par M.Ni pas comprise*)

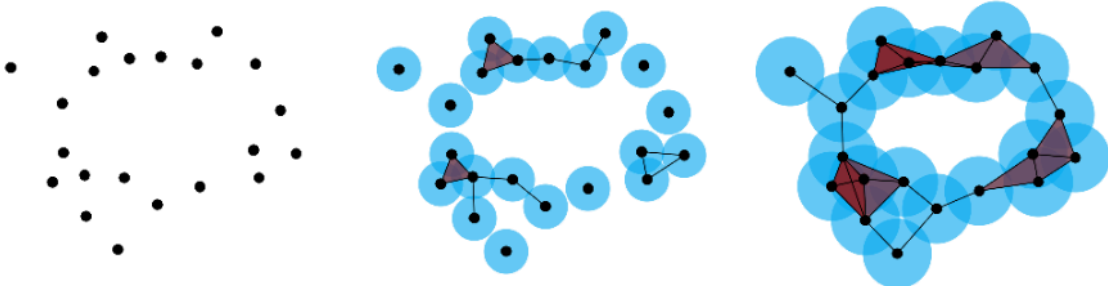


Figure 5: Construction d'un complexe simplicial avec un r grandissant de gauche à droite

Ici, r est le rayon des boules bleues, et un simplexe est considéré dès lors que les boules associées à ces sommets se rencontrent.

Pour définir formellement des “trous”, nous devons définir les opérateurs de bords :

Définition :

On définit un complexe de chaînes comme la donnée d'une suite

$$\dots \xrightarrow{\delta_{k+2}} C_{k+1} \xrightarrow{\delta_{k+1}} C_k \xrightarrow{\delta_k} C_{k-1} \xrightarrow{\delta_{k-1}} \dots$$

Où chaque C_k est un groupe abélien et δ_k est une morphisme de groupes tel que $\delta_k \circ \delta_{k+1} = 0$

On appelle δ_k un *opérateur de bords*.

Définition :

On définit alors les *classes d'homologie de dimension k* comme le groupe de $\text{Ker}(\delta_k)$ quotienté par $\text{Im}(\delta_{k+1})$:

$$H_k = \text{Ker}(\delta_k) / \text{Im}(\delta_{k+1})$$

Celle ci représente les “trous” en dimension k

On peut voir que H_0 représente des points connectés, H_1 représente un trou qui est entouré par un chemin fermé de points connectés (K_2 dans Figure 3 par exemple) et H_2 représenterait par exemple une pyramide d'intérieur vide.

Pour notre usage, nous voulons calculer H_0 les temps moyens pour se rendre à une station de métros et H_1 qui représente les zones critiques de couverture du réseau.

II Méthode

Pour trouver les zones critiques, nous pouvons appliquer la méthode suivante décrite dans [2]. Celle ci se décompose en 4 étapes :

- Récupération de l'ensemble des stations, leurs poids, et les distances entre elles;
- Transformation de ces informations en une filtration;
- Création et réduction de la matrice de bordure (définie dans la suite);
- Récupération des simplexes “tueurs” de classes d'homologies

La première étape étant développée en section III, nous supposons dans la suite de cette section avoir un espace muni d'une distance $((x_i)_i, d)$ et où chaque x_i admet w_i pour pondération.

À l'étape 2, nous allons créer une filtration grâce à la définition donnée section I.

A partir d'une filtration donnée, nous pouvons obtenir les classes d'homologie grâce au théorème qui suit :

Théorème :

D'après [1] et [3], chaque H_k est isomorphe à une somme directe d'espaces filtrés associés à une certaine famille d'intervalles, définie d'une unique manière. Cette famille d'intervalles est appelé un *code barre*.

Informatiquement, selon [4], on calcule ce code barre en créant une matrice de bordure B après avoir défini un ordre total sur les simplexes respectant les propriétés énoncées section I

Définition :

On définit la matrice de bordure, associée à un ordre total $\sigma_0 < \dots < \sigma_{n-1}$ sur tous les simplexes $(\sigma_i)_{i=0}^{n-1}$ de la filtration, suivant :

$$\forall (i, j) \in \llbracket 0, n-1 \rrbracket^2, B[i][j] = \begin{cases} 1 & \text{si } \sigma_i \text{ est une face de } \sigma_j \\ 0 & \text{sinon} \end{cases}$$

Note : D'après [3], B peut être vu comme la matrice de δ_1 dans la base canonique associés aux complexes de chaînes.

Un exemple d'une telle matrice est donnée en Table 1.

Après avoir calculé B , nous voulons la *réduire* en code barre, dans le sens où l'on peut interpréter correctement les valeurs de cette matrice avec la filtration (grâce au théorème précédent). Le terme de *réduction* fait ici référence au fait qu'à partir de B , nous allons vouloir attribuer à chaque simplexe la naissance d'*au plus* une classe d'homologie, nous pouvons observer ce résultat en Table 2.

Cet algorithme de réduction est nommé *standard algorithm* et est décrit dans [4] par, en posant $\text{low}_B(j) = \max(\{i \in \llbracket 0, n-1 \rrbracket, B[i][j] \neq 0\}) \in \mathbb{N} \cup \{-1\}$:

StandardAlgorithm(B)

```
for j in  $\llbracket 0, n-1 \rrbracket$ :
    while (il existe i < j avec  $\text{low}_B(i) = \text{low}_B(j)$ ):
        ajouter colonne i de B à colonne j
```

Comparons alors nos deux matrices, sur l'exemple de la filtration de Figure 3 (Les cases vident remplacent les zeros pour plus de lisibilité), avec ici notre ordre total sur les simplexes :

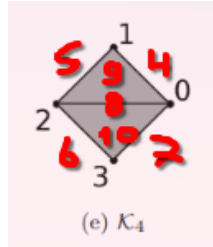


Figure 6: Nommage des simplexes (A rendre plus beau)

En regardant la matrice \overline{B} , nous remarquons que l'opération de réduction à permis d'avoir la ligne low sans répétition de nombre positifs, autrement dit, on accorde la naissance d'un simplexe à un unique autre simplexe. Ainsi le simplexe 1 donne naissance au simplexe 4, 8 donne naissance à 9 et 7 donne naissance à 10.

Par exemple, si $\text{low}_{\overline{B}}(j) = i \neq -1$ alors on a une paire de simplexe (σ_i, σ_j) telle que l'apparition de σ_i fait apparaître une nouvelle classe d'homologie. Et au contraire, σ_j va la *tuer* en apparaissant. Prenons comme exemple la filtration Figure 3 : dans K_0 , le point 1 cause l'apparition d'une classe dans H_0 cependant l'apparition du simplexe $(0,3)$ dans K_2 tue la classe de 1 dans H_0 mais créer une nouvelle classe dans H_1 (car elle crée un cycle).

	0	1	2	3	4	5	6	7	8	9	10
0					1			1	1		
1					1	1					
2						1	1		1		
3							1	1			
4										1	
5										1	
6											1
7											1
8										1	1
9											
10											
low	-1	-1	-1	-1	1	2	3	3	2	8	8

Table 1: Matrice B non réduite

	0	1	2	3	4	5	6	7	8	9	10
0					1						
1					1	1					
2						1	1				
3							1				
4										1	1
5										1	1
6											1
7											1
8										1	
9											
10											
low	-1	-1	-1	-1	1	2	3	-1	-1	8	7

Table 2: Matrice \overline{B} , la matrice B après réduction

En revanche si $\text{low}_{\overline{B}}(j) = -1$ alors l'apparition de σ_j crée une classe d'homologie : s'il existe k tel que $\text{low}_{\overline{B}}(k) = j$ on est dans le cas précédent, sinon la classe d'homologie n'est jamais tuée.

C'est depuis cette matrice que nous sommes capables de déterminer H_0 ainsi que H_1 , et donc de générer des représentations graphiques comme montré en Figure 8

III Les données

III.1 Sources

Ne voulant pas me baser sur des villes factices, j'ai décidé de trouver des sources pouvant me fournir des informations sur les stations de metros de plusieurs grandes villes de France comme Toulouse ou Marseille.

Ainsi, toutes les informations relatives aux stations de metros ainsi que les temps de passages sont trouvable sur le site du gouvernement : <https://transport.data.gouv.fr>.

Ces informations servent à définir nos points et notre pondération (voir Section III.2), en revanche elles ne permettent pas d'obtenir les distances entre les stations, pour cela nous utiliserons alors <https://www.geoapify.com> qui nous permet d'estimer des temps de trajet en voiture et à pied.

III.2 Construction des informations importantes

Définissons dès lors nos objets :

Définition :

Un point x_i , représentant une station de métro, est défini par deux données, celle de la position géographique (latitude/longitude) ainsi que son poids w_i . Le poids w_i est égal à la moyenne du temps d'attente entre deux métros en station x_i sur une semaine entière.

Les temps de passage des metros en station étant plus ou moins constant sur la semaine, il est cohérent d'utiliser une moyenne.

On définit la distance similairement à [2]:

Définition :

On définit la distance entre deux stations de métros x et y comme :

$$d(x, y) = \frac{1}{2}(\min(t_{\text{marche}}(x, y), t_{\text{voiture}}(x, y)) + \min(t_{\text{marche}}(y, x), t_{\text{voiture}}(y, x)))$$

Ainsi en revenant aux boules des complexes simplicaux de Vietoris-Rips, elle modélise le coût temporel d'un trajet "porte à porte" en utilisant le métro.

IV Résultats et conclusion

Ville	Dimension	Médiane	Variance
Toulouse	0D Homologie	211.00s	22.27s
	1D Homologie	318.00s	58.62s
Marseille	0D Homologie	184.00s	23.57s
	1D Homologie	223.50s	10.5s

Table 3: Tableau récapitulant les médianes ainsi que la variance des temps de mort des classes holomogiques pour chaque ville.

On comprend que globalement il faut 200s (soit 3m20 environ) pour quelqu'un de se rendre d'une station à une autre (le minimum en temps entre la voiture et la marche) ce qui est effectivement cohérent avec la réalité. Les temps des classes 1D ici présent montre le temps moyen de trajet entre les deux stations les plus éloignées d'un même cycle. Donc par exemple pour Toulouse, il faudra en moyenne 318s (5min20s) pour rejoindre une station depuis les zones les moins biens deservies.

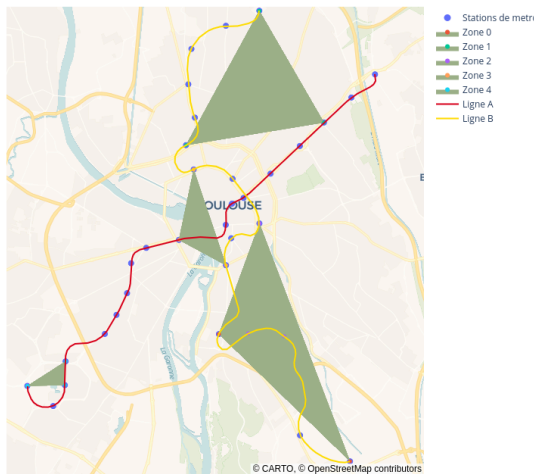


Figure 7: Carte de toulouse

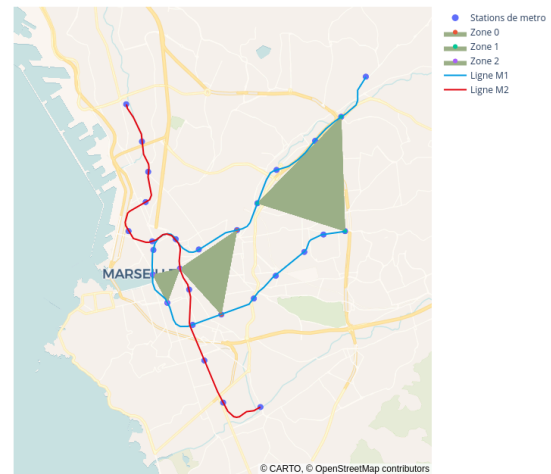


Figure 8: Carte de Marseille

Les triangles ici représentés montrent les zones où il est le plus difficile de rejoindre une station de métro. Pour les triangles les plus gros, il peut être cohérent de croire qu'il est difficile de se rendre à ces stations de métros. En revanche, pour les plus petits cela est plus dur.

Ce sont des zones où il ne circule que très peu de voitures entre les stations de métros, en effet ces zones sont uniquement piétonnes donc la distance parcourue à durée égale est nécessairement plus long à pied qu'en voiture. Donc la distance prise par notre algorithme est celle relevant de la marche à pied, d'où les zones *a priori* plus petites que celles discutées plus haut.

Nous pouvons observer le plus gros problème de cette méthode : la méthode est pertinente pour le développement d'un réseau autre que le métropolitain (réseau de bus, par exemple). En effet, on remarque que les zones critiques sont entre les lignes de metros dessinées, mais jamais en bout de ligne, là où pourtant la disponibilité des métros est plus faible que dans l'hypercentre des villes.

Mais via cette interprétation là, je n'arrive pas à expliquer la présence du triangle au sud de Toulouse, parce qu'il est en bout de ligne, même si effectivement le temps de trajet est haut dans cette région de la ville.

Ainsi, cette méthode d'analyse peut être pertinente lors d'une simulation pour la création ou l'amélioration prévue d'un réseau, afin de détecter les zones qui seront le plus en besoin avec le réseau imaginé, mais ne permet pas d'établir un tracé *optimal* d'une ligne de métro pour satisfaire le plus de monde.

Bibliography

- [1] Henri Paul de Saint-Gervais, Une invitation à l'homologie persistante, (n.d.).
- [2] ABIGAIL HICKOK, BENJAMIN JARMAN, MICHAEL JOHNSON, JIAJIE LUO, and MASON A. PORTER, *Persitent Homology for Resource Coverage: A Case Study of Access to Polling Sites*.
- [3] Afra Zomorodian and Gunnar Carlsson, *Computing Persistent Homology*.
- [4] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington, *A Roadmap for the Computation of Persistent Homology*.