

Appareiller des données géographiques

Pavard A.^{1,2}

1. Département Génie de la construction, École de Technologie Supérieure, Montréal, QC, H3C 1K3, Canada.
2. Groupe de Recherche en Intégration et Développement Durable en Environnement Bâti, Montréal, QC, H3C 1K3, Canada

antonin.pavard.1@ens.etsmtl.ca

Cet article a été présenté à l'occasion du 24ème Salon Culture Jeux & Mathématiques qui s'est déroulé du 25 au 28 Mai 2023 (<https://salon-math.fr/>). Il s'agit d'un article de vulgarisation qui s'appuie sur des travaux réalisés dans le cadre de la thèse de Doctorat d'Antonin Pavard sur la conception d'un référentiel spatial de voirie pour l'optimisation de la gestion de l'infrastructure dans son environnement. La méthode complète et les résultats sont diffusés dans le manuscrit de thèse (Pavard, 2020a) et ont été présentés lors d'une soutenance filmée (Pavard, 2020b). Plus précisément, cet article reprend une étape du processus de conception du référentiel : la construction de données géographiques.

Le monde réel est composé d'objets géographiques tels que les villes ou la voirie. Ces objets sont modélisés et représentés dans différentes bases de données (BD) géographiques pour répondre à divers besoins. Par exemple, il existe des BD de voirie construites pour représenter les axes de circulation ou pour calculer les temps de trajet, et d'autres BD pour représenter les emprises d'infrastructures sur l'espace (Pavard, 2023). Parfois, il est nécessaire de relier différentes BD représentant des objets géographiques identiques. Pourquoi et comment relier ces bases de données géographiques ?

Relier des bases de données géographiques : pourquoi ?

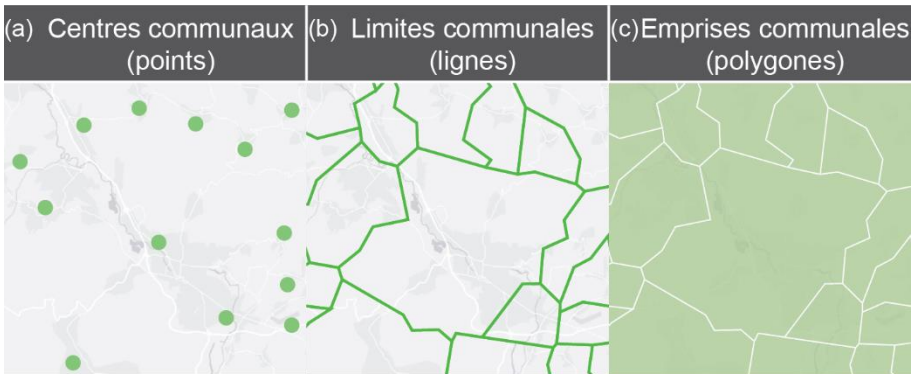
En science de l'information géographique, les systèmes d'information géographique (SIG) permettent de représenter les objets géographiques dans leur environnement à l'aide de deux composantes : *spatiale*, et *descriptive*.

La *composante spatiale* correspond à la géométrie. Elle donne une information sur un objet géographique : sa position dans l'espace, son étendue, et sa topologie (c'est-à-dire les relations spatiales entretenues avec d'autres objets).

Dans un SIG, trois *objets primitifs* sont utilisés pour reconstituer la géométrie d'un objet géographique : le *point*, la *ligne*, le *polygone*. Le choix de ces objets primitifs dépend des usages de la BD produite.

La figure suivante illustre le cas d'une BD *communale*, où l'objet « commune » est représenté dans trois *couches géographiques* différentes (Fig. 1) :

- ✓ par un point pour sa localisation (a). On parle alors de données *ponctuelles* ;
- ✓ par une suite associée de segments de ligne, une *polyligne*, pour ses limites géographiques (b). On parle dans ce cas de données *linéaires* ;
- ✓ par un polygone pour son emprise (c). On parle de données *surfaiques*.



Sources : IGN, GEOFLA®, 2013

Figure 1. Trois exemple de représentation d'une commune dans les SIG (Pavard, 2022)

La *composante descriptive* donne des informations thématiques sur les objets géographiques à l'aide d'*attributs alphanumériques*. Ces informations sont structurées et organisées dans des *tables attributaires*. La figure suivante illustre ce point dans le cas de la BD communale (Fig. 2). Par exemple :

- ✓ deux attributs identiques sont associés aux objets des couches géographiques (ici, un identifiant et le nom des communes) ;
- ✓ des attributs distincts apportent des informations supplémentaires sur les objets géographiques selon les couches, telles que la population pour la couche points, la longueur du périmètre pour la couche lignes et la superficie pour la couche polygones.

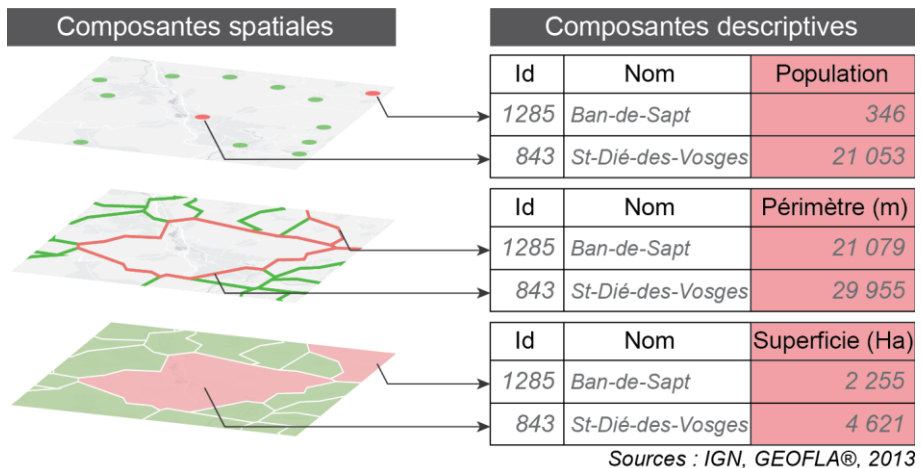


Figure 2. Exemples de tables attributaires pour décrire des communes (Pavard, 2022)

L'association de la composante spatiale à la composante descriptive est une force des SIG. Grâce à elle, il est par exemple possible de sélectionner en un « click de souris » toutes les villes se trouvant dans un certain rayon de la capitale mondiale de la géographie, qui se trouve à Saint-Dié-des-Vosges. Quant aux informations descriptives, elles permettent d'analyser le territoire et d'en comprendre certains enjeux qui peuvent y être associés (en croisant des données sur les populations et sur les emplois par exemple, ou encore en calculant des densités démographiques à l'aide des superficies et des populations). De telles analyses nécessitent d'associer des données contenues sur plusieurs couches de données.

Dans l'exemple donné ici, le transfert d'une table de données à l'autre se réalise simplement car la composante spatiale est reliée par un *identifiant unique* : 843 pour Saint-Dié-des-Vosges. On dit alors que les données sont *appariées*. Ici, l'appariement est assuré dès la phase de construction des données : les couches constituant la BD communale Geofla (IGN, 1997), devenue Admin Express en 2016, sont produites par un organisme unique, l'Institut de Géographie et forestière (IGN). Dans d'autres cas l'appariement est réalisé a posteriori. Des recherches portent par exemple sur le cas de (Fig. 3) :

- ✓ deux couches surfaciques (a) lors d'analyse sur les villes. Différentes définitions des villes donnent des emprises différentes (Bretagnolle et al., 2014 ; Pumain et al., 1992) ;
- ✓ deux couches linéaires (b) lors d'analyse sur réseaux (Mustière, 2006 ; Devogèle, 1997). Un même réseau de voirie peut être représenté à l'aide de graphes ayant des *généralisations* différentes, soit des différences au niveau de la *sélection* et de la *schématisation* des objets à représenter ;
- ✓ une couche linéaire et une couche surfacique (c) lors d'analyses sur des objets géographiques de type réseau, nécessitant de prendre en compte l'emprise. **Intéressons-nous ici à ce cas, qui est moins traité en recherche.**

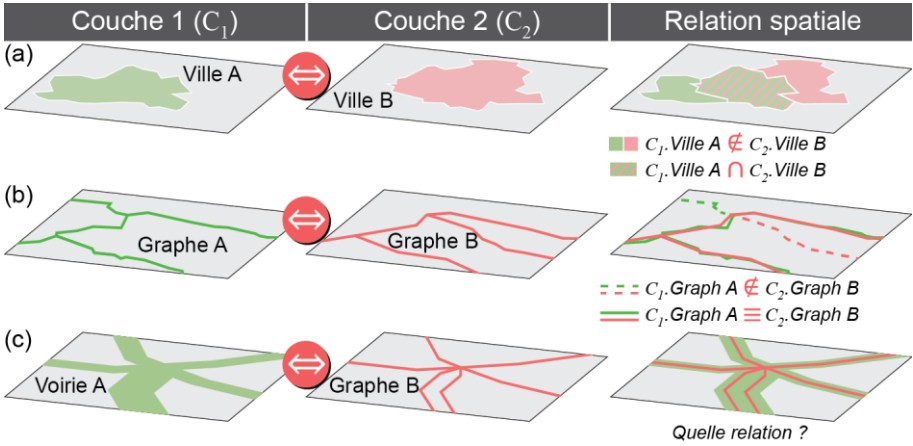


Figure 3. Appariement de données géographiques (Pavard, 2022)

L'appariement : des mathématiques aux SIG

Pour les questions d'appariement, les mathématiques sont essentielles. Elles permettent d'expliquer les relations entre deux ensembles de données : l'ensemble de départ et l'ensemble d'arrivée. Ici l'ensemble de départ $E.d$ correspond à nos données lignes, et l'ensemble d'arrivée $E.a$ à nos données polygones. Chaque ensemble comporte un à plusieurs éléments, tels que des lignes ou des polygones. Les relations sont des *applications* lorsque tous les éléments de l'ensemble de départ sont liés aux éléments de l'ensemble d'arrivée. Les trois applications les plus courantes en mathématiques sont (Fig. 4) (Armstrong, 2013) :

- ✓ la *bijective* (1-1) lorsque chaque élément de $E.a$ correspond à un seul élément de $E.d$ (a) ;
- ✓ la *surjective* (n-1 ou 1-n) lorsque chaque élément de $E.a$ correspond à au moins un élément de $E.d$ (b) ;
- ✓ la *multivaluée* (n-p) lorsque chaque élément d'un ensemble correspond à plusieurs éléments de l'autre ensemble (c).

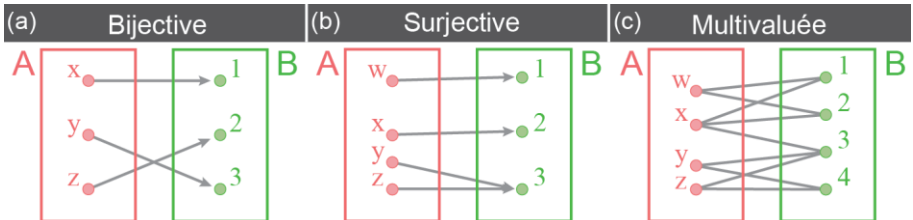


Figure 4. Les applications courantes en mathématiques (adaptée de Pavard, 2020a)

Dans la figure suivante, ces applications sont transcrites pour décrire des relations spatiales de type « same as » (Clementini and Laurini, 2008) les plus typiques entre des objets primitifs lignes et polygones dans le cas de données géographiques de type réseau (Fig. 5) :

- ✓ le type 1-1 (a) est une relation forte et permet une association spatiale facile ;
- ✓ le type 1-n ou n-1 (b) est une relation moyenne dont l'association spatiale nécessite une opération simple ;
- ✓ le type n-p (c) est une relation faible dont l'association spatiale est la plus difficile et nécessite une opération complexe.

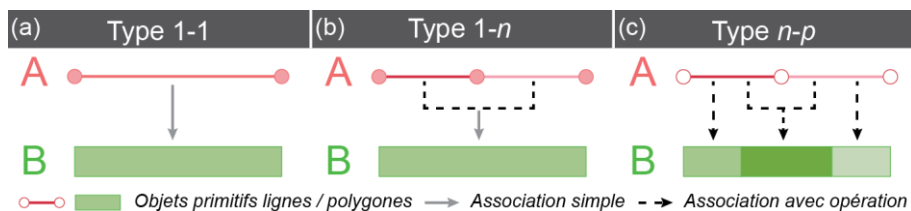
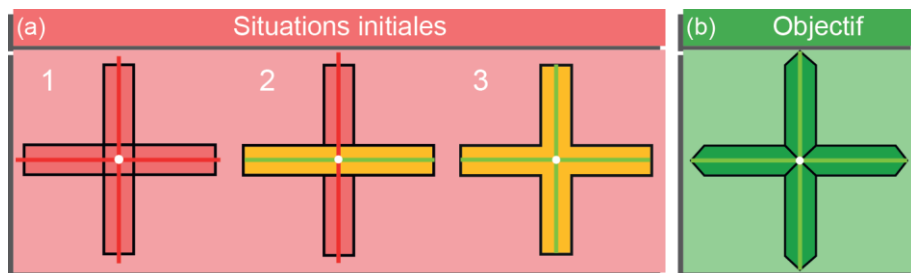


Figure 5. Relations spatiales typiques (Pavard, 2022)

L'appariement, de données linéaires et surfaciques : comment ?

La voirie est un bon exemple pour illustrer l'appariement de données linéaires de type graphe et surfaciques. D'abord, rappelons que la voirie est composée de *sections courantes* reliant les extrémités des impasses et de carrefours. L'appariement entre les objets primitifs lignes et polygones est alors surtout importante au niveau des carrefours. En effet, on y observe les relations de type n-p. Pour simplifier, appuyons-nous sur le cas suivant : le carrefour « en croix », soit lorsque quatre sections courantes débouchent sur une intersection. Les graphes de voirie sont toujours structurés de la même façon : des arcs représentant les sections courantes relient des nœuds représentant des carrefours ou des extrémités d'impasse (Fig. 6).



Relation spatiale

Objet polygone : ■ Relation n-p ■ Relation 1-n ■ Relation 1-1

Objet ligne : — Relation n-p — Relation 1-1

Figure 6. Structuration des graphes de voirie (Pavard, 2022)

Les données surfaciques peuvent varier. Ainsi, elles sont construites classiquement (a) soit en modélisant distinctement les carrefour (1), soit en associant un carrefour à deux sections courantes (2), soit en agréant l'ensemble (3) (Pavard et al., 2022b). Les relations spatiales sont au mieux « moyennes ». De façon à simplifier les relations spatiales et assurer l'appariement, il est possible de structurer les objets surfaciques de sorte qu'un objet polygone ne soit lié qu'à un seul objet linéaire (b).

Notre solution (Pavard et al., 2022a) consiste à diviser les polygones aux carrefours en autant de portions que de sections courantes qui y débouchent. Pour cela, on utilise la *tessellation*. Cette technique mathématique consiste à couvrir la totalité d'un espace avec un nombre fini de polygones non chevauchants construits autour de points sources. En d'autres termes, la *tessellation* est une partition, qui divise un territoire donné en autant de régions (polygones) que de points sources.

Chacune de ces divisions, appelées *cellules de Voronoï*, englobe un unique point source (Dupuis, 2003). Une cellule de Voronoï représente l'ensemble des points du plu « les plus proches » (au sens de la distance euclidienne usuelle) de ce point source. Voici comment construire ces cellules (Fig. 7).

On part d'une situation initiale (a), à savoir des points répartis sur un espace. Les points sont reliés entre eux de sorte à produire une triangulation (b). Elle est Delaunay lorsque aucun autre sommet que ceux participant à un triangle ne se trouve dans le cercle inscrit de ce triangle. Par exemple, les points P_2 et P_4 ne peuvent être reliés pour former le triangle $P_1 P_2 P_4$ sans quoi le sommet P_3 se trouverait dans le cercle circonscrit de ce triangle. Ensuite, les milieux de chaque segment des triangles sont repérés (c). Pour finir, on construit le diagramme de Voronoï (d) de sorte que chaque cellule n'englobe qu'un seul point. Pour cela, une cellule est obtenue en reliant les milieux de tous les segments partant d'un point donné.

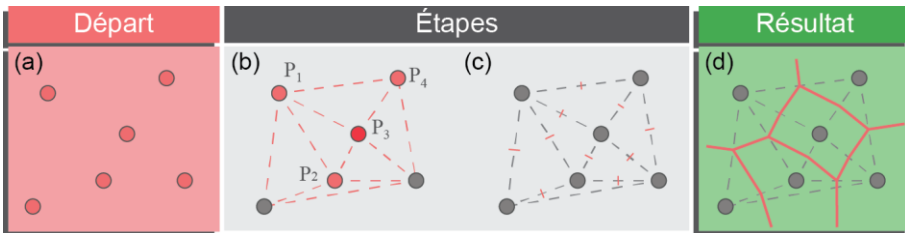
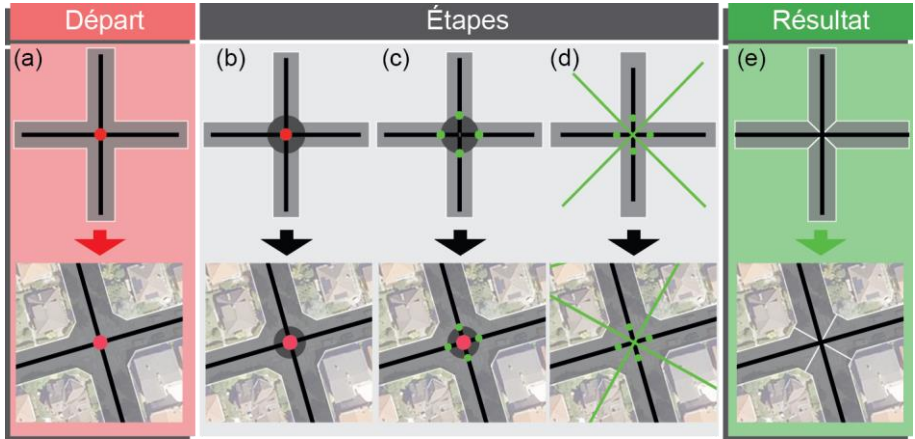


Figure 7. Triangulation de Delaunay cellules de Voronoï (Pavard, 2022)

La figure ci-après illustre les étapes à travers un cas abstrait d'un carrefour puis son application sur des données décrivant un territoire réel :

- ✓ au départ, on dispose de deux jeux de données représentant un carrefour (a) : un jeu linéaire et un surfacique ;
- ✓ la *tessellation* construit une cellule pour un point source. Ainsi, partir d'un seul point situé au centre du carrefour conduit à englober celui-ci dans une seule cellule. L'objectif est de diviser le carrefour en autant de portions que de sections courantes. On construit alors une zone tampon ayant un diamètre supérieur à la largeur d'une chaussée, de façon à englober le carrefour (b) ;

- ✓ les points sources sont construits à l'intersection de la zone tampon et des objets linéaires représentant le carrefour (c) ;
- ✓ les points sources servent à construire les cellules de *tessellation* dont les limites s'établissent à égale distance de chaque point source (d) ;
- ✓ finalement, les données surfaciques de voirie sont fractionnées à l'aide des cellules de *tessellation* (e).



Sources : IGN, BD TOPO®, 2016 ; BD ORTHO®, 2019

Figure 8. Illustration de la démarche sur un carrefour (Pavard, 2022)

Cette méthode appliquée sur l'ensemble des données de voirie d'un territoire permet d'appareiller une couche linéaire et une couche surfacique décrivant une même voirie. À partir de là, il est possible de transférer des données du linéaire vers le surfacique (et inversement). Des informations de trafic traditionnellement cartographiées sur des objets linéaires sont associées aux objets surfaciques de façon à réaliser par exemple des vérifications de dimensionnement (Pavard, 2023). La méthode d'appariement actuelle, doit encore être améliorée de façon à traiter les configurations les plus élaborées telles que les échangeurs autoroutiers.

Références

- Armstrong, M. A. (2013) Basic Topology. *Springer Science & Business Media*, New Delhi
- Bretagnolle, A., Guérois, M., Mathian, H., & Pavard, A. (2014). *UMZ: a database now operational for urban studies (M4D improvements)* (Doctoral dissertation, ESPON).
- Clementini, E. et Laurini, R. (2008) Un cadre conceptuel pour modéliser les relations spatiales. *Revue des Nouvelles Technologies de l'Information (RNTI) RNTI-E-14*, 1–17.
- Devogèle, T. (1997) Processus d'intégration et d'appariement de Bases de Données Géographiques; Application à une base de données routières multi-échelles. PhD thesis. Université de Versailles-Saint Quentin en Yvelines.
- Dupuis, F. (2003) Tessellations de Voronoï appliquées aux structures protéiques. PhD thesis. Université Paris-Diderot - Paris VII.
- Mustière, S. (2006) Results of experiments on automated matching of networks at different scales. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **36**, 92–100
- Pavard, A. (2023). La gestion de la voirie. Zenodo. <https://doi.org/10.5281/zenodo.7974430>
- Pavard, A., Bordin, P. Dony, A. (2022a). « Coherence of linear and surface data: Methodological proposal from the example of road data », arXiv preprint, arXiv:2209.06803 ; DOI : <https://doi.org/10.48550/arXiv.2209.06803>.
- Pavard A., Bordin P., Dony A. (2022b) Modéliser la voirie urbaine pour une gestion de l'infrastructure routière : étude comparée de deux conceptions de données surfaciques de voirie, *Belgeo* [En ligne], **3** | 2021, mis en ligne le 25 janvier 2022, consulté le 21 janvier 2023. URL : <http://journals.openedition.org/belgeo/51958> ; DOI : <https://doi.org/10.4000/belgeo.51958>
- Pavard A. (2020a) Optimiser la gestion conjointe de la voirie et des réseaux enterrés à l'aide de la géomatique : conception d'un référentiel spatial de voirie – Constructibilité à l'interface entre route et réseaux enterrés. Géographie. Université Paris-Est. Français. (NNT : 2020PESC2066).
- Pavard A. (2020b) Soutenance de Doctorat : Optimiser la gestion conjointe de la voirie et des réseaux enterrés à l'aide de la géomatique : conception d'un référentiel spatial de voirie. (hal-03346851)
- Pumain, D., Saint-Julien, T., Cattant, N. et Rozenblat C. (1992) The statistical concept of the town in Europe. Office for Official Publication of the European

Communities, Luxembourg.