



Enrichissement automatique de données par analyse spatiale pour la généralisation de réseaux

Guillaume Touya

► To cite this version:

Guillaume Touya. Enrichissement automatique de données par analyse spatiale pour la généralisation de réseaux. *Revue Internationale de Géomatique*, 2010, 20 (2), pp.175-199. 10.3166/rig.20.175-199 . hal-02309189

HAL Id: hal-02309189

<https://hal.science/hal-02309189>

Submitted on 9 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enrichissement automatique de données par analyse spatiale pour la généralisation de réseaux

Guillaume Touya

Laboratoire COGIT, Institut Géographique National
2 Avenue Pasteur 94160 Saint-Mandé Cedex
{ prénom.nom}@ign.fr

RÉSUMÉ. La généralisation est un processus qui vise à réduire le niveau de détail d'une base de données géographiques dans le but de satisfaire de nouvelles spécifications. La sélection est une étape importante de ce processus qui consiste à choisir en fonction du contexte géographique et des besoins les objets qui devraient apparaître dans la base de données généralisée. Cet article présente une méthode générale pour la sélection automatique de réseaux géographiques basée sur l'enrichissement de données par analyse spatiale. La base de données géographiques initiale est enrichie par la reconnaissance d'informations, de structures qui n'étaient pas explicites. Par exemple, dans un réseau routier, une aire d'autoroute est un ensemble de tronçons de route connectés entrant et sortant de l'autoroute. Les processus de sélection, plus spécifiques à chaque thème, prennent en compte cet enrichissement et traitent de manière adaptée ces structures qui sont essentielles pour les futures applications de la base de données généralisée (analyse, cartographie). La méthode présentée est mise en oeuvre pour les réseaux routier et hydrographique puis testée sur des données topographiques issues des bases de données de l'Institut Géographique National. L'intérêt d'un tel enrichissement pour d'autres objectifs que la généralisation est étudié.

ABSTRACT. Generalisation is a process that seeks to reduce the level of detail of a geographic database in order to meet new specifications. Selection is a major step of the process that consist in choosing the objects that will be part of the generalised database according to geographical context. This paper presents a general method for automated selection of geographical networks based on data enrichment by spatial analysis. The initial database is enriched with implicit information or structure recognition. For instance, in a road network, highway services are groups of connected roads with enter and exit to the highway. The selection process, specific for the different themes, takes the enriched data into account and handle such structures that are essential for the future applications of the database (analysis, cartography) with appropriate treatment. The presented method is carried out on road and river networks and applied to topographic data coming from IGN databases. The interest of such enrichments for other application is also considered.

MOTS-CLÉS : généralisation, sélection, réseaux, analyse spatiale, graphes.

KEYWORDS: generalisation, selection, network, spatial analysis, graphs.

1. Introduction

La généralisation est un processus qui vise à réduire le niveau de détail d'une base de données géographiques dans le but de satisfaire de nouvelles spécifications. Lorsque l'objectif est de produire une carte à une certaine échelle, on parle de généralisation cartographique. Lorsque l'objectif est la dérivation d'une nouvelle base de données géographiques, on parle de généralisation de modèle ou de base de données (Weibel et Dutton, 1998). La généralisation de base de données peut aussi être un préalable à une généralisation cartographique. La sélection est une étape clé du processus de généralisation de base de données : elle consiste à choisir parmi les objets géographiques de la base de données initiale ceux qui feront partie de la base de données généralisée car répondant aux spécifications de cette dernière. Pour savoir si un objet répond aux spécifications, il est souvent nécessaire de faire appel au contexte de l'objet et à des concepts géographiques implicites, ce qui rend ce processus complexe. Cet article présente une méthode générale pour la sélection des réseaux géographiques basée sur l'enrichissement de données par analyse spatiale. Cette méthode est mise en oeuvre pour la sélection des réseaux routier et hydrographique.

L'enrichissement de données consiste à détecter et expliciter dans une base de données des informations, des concepts ou des structures géographiques implicites. Par exemple, sans que l'information soit explicite dans les données, un humain peut, en visualisant un réseau routier, reconnaître une aire d'autoroute (Figure 1) : c'est un ensemble de tronçons de routes connectés entre eux dont l'un permet de sortir de l'autoroute et un autre d'y rentrer. Dans ce cas particulier, en se basant sur nos connaissances géographiques, l'enrichissement consiste à détecter automatiquement ce groupe de tronçons, à le qualifier et à l'intégrer à la base de données.

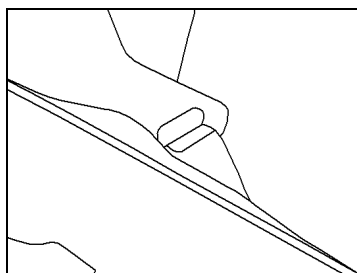


Figure 1. Une aire d'autoroute dans un réseau routier : un ensemble de tronçons dont la structure fait sens.

L'enrichissement de données par analyse spatiale est jugé comme primordial pour la généralisation par de nombreux auteurs (McMaster et Shea, 1988, Weibel et Dutton, 1998, Mackaness et Edwards, 2002). L'enrichissement dans les réseaux a

d'ailleurs été étudié, particulièrement pour les routes (Heinzle et Anders, 2007). Ces travaux font apparaître que les réseaux géographiques contiennent des structures importantes pour la compréhension de ces données géographiques telle qu'un échangeur routier ou un boulevard périphérique dans un réseau routier ou un delta dans un réseau hydrographique. Différents travaux de recherche traitent de la sélection du réseau routier (Richardson et Thomson, 1996, Thomson et Richardson 1999) et du réseau hydrographique (Horton, 1945, Thomson et Brooks, 2000) mais ne prennent pas en compte toutes ces informations et structures implicites.

L'objectif de ce travail est d'une part de mettre au point des méthodes d'analyse spatiale pour expliciter le contexte spatial dans les réseaux et détecter les structures caractéristiques de ces réseaux. Pour cela, nous adaptons des méthodes issues de la littérature et proposons de nouvelles méthodes d'analyse spatiale. D'autre part, l'objectif est de concevoir une méthode générale pour la sélection des réseaux se basant sur cet enrichissement par analyse spatiale pour mieux traiter le contexte géographique des réseaux et pour maintenir durant le processus les structures importantes des réseaux. Le but est également de mettre en oeuvre la méthode pour des processus de sélection des réseaux routier et hydrographique.

La deuxième partie de cet article présente la méthode générale proposée de sélection des réseaux. La troisième décrit les enrichissements réalisés en deux temps: la détection des structures importantes et leur caractérisation. La quatrième partie présente les processus de sélection mis en oeuvre pour les réseaux routiers et hydrographiques avec une application sur des données topographiques issues des bases de données de l'Institut Géographique National (IGN). La cinquième partie explore l'utilisation des enrichissements à d'autres fins que la généralisation. Enfin, la dernière partie dresse le bilan de ce travail et expose quelques perspectives de recherche.

2. Méthodologie proposée pour la sélection des réseaux

Afin de bien comprendre et de résoudre au mieux tous les problèmes posés par la sélection de réseaux lorsque l'on souhaite tenir compte des structures caractéristiques, nous avons implémenté plusieurs méthodes issues de la littérature et effectué des tests avec ces méthodes sur un jeu de données test issu de la BD TOPO® de l'IGN (résolution métrique). Ces tests ont notamment mis en valeur l'importance de l'enrichissement des données mais aussi révélé la nécessité d'appliquer des traitements de typification spécifiques à certaines structures plutôt que de les intégrer à un processus traitant tous les éléments de réseau de la même manière. Ces tests ont également permis d'identifier les points qui nécessitent de développer de nouvelles méthodes que ce soit pour l'enrichissement ou la sélection. L'analyse de ces tests a donc abouti à la conception d'une méthode générale pour la sélection de réseaux géographiques, que nous présentons en Figure 2.

La première étape de la méthode proposée est l'enrichissement de la base de données par les informations et structures implicites nécessaires pour répondre aux spécifications de la base de données généralisée. En reprenant, les définitions de (Ruas, 1999, p.111), nous appelons structure toute catégorie d'objet que l'on retrouve dans les données composé de plusieurs objets simples comme une ville (composée de ses routes et ses bâtiments) ainsi que toute configuration géographique particulière comme un alignement de maisons ou dans notre cas une aire d'autoroute. Nous appelons structure caractéristique une structure dont la configuration particulière est telle qu'elle ne pourrait pas être conservée par une sélection standard. Un processus de sélection standard traite les tronçons d'un réseau de manière uniforme pour déterminer lesquels doivent être conservés. Un échangeur routier est un exemple de structure caractéristique du réseau routier. Un processus de sélection standard n'est pas adapté aux échangeurs routiers. Ainsi, si elle n'est pas identifiée (c'est-à-dire détectée) au préalable, des composants de la structure comme des bretelles risquent de ne pas être sélectionnés, l'échangeur routier n'étant plus reconnaissable dans les données alors qu'il présenterait pourtant un intérêt dans la base de données généralisée. Dans notre processus générique de sélection, les structures sont caractérisées pour être idéalement utilisées par les processus suivants dans cette approche générale. Les données enrichies servent d'entrée au processus de sélection standard mais également de contrôle car certains des enrichissements guident le processus en explicitant le contexte géographique des différents objets. Par exemple, la présence de méandres sur une rivière va guider la sélection de manière à préserver ces méandres.

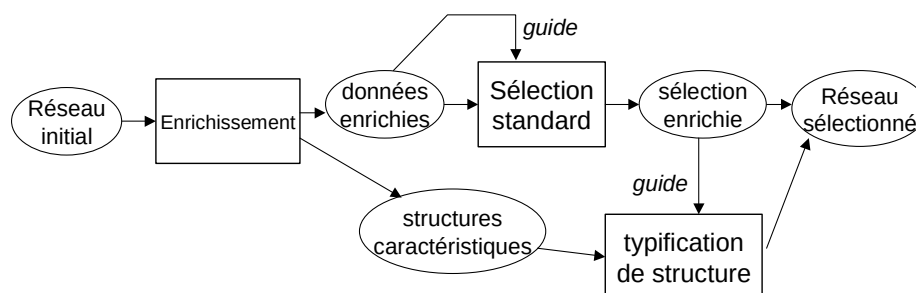


Figure 2. Schéma de la méthode générale de sélection des réseaux basée sur l'enrichissement de données dans un formalisme de type SADT.

L'étape suivante est alors le processus de sélection standard qui, guidé par les enrichissements, va travailler sur les données enrichies pour choisir celles qui feront partie du réseau sélectionné. Les données enrichies sont les données initiales plus les structures non caractéristiques détectées et caractérisées moins les structures caractéristiques détectées. Cette partie est plus spécifique à chaque réseau : par exemple, pour le réseau routier, la sélection est décomposée en trois étapes, la sélection en zone rurale, la sélection en zone urbaine et la vérification de la

continuité à l'interface des deux zones. Enfin, la dernière étape de la méthode générale est la typification des structures caractéristiques du réseau. Il s'agit de diminuer le niveau de détail de la structure tout en maintenant ou en caricaturant ses caractéristiques principales. Cette étape est principalement effectuée par des algorithmes ad-hoc qui sont guidés par les résultats de la sélection sur le réseau général. Par exemple, dans le cas d'une zone d'irrigation en forme de quadrillage dans un réseau hydrographique, la typification consiste en une réduction du nombre de mailles tout en conservant la forme d'un quadrillage. Ces étapes d'enrichissement, de sélection standard et de typification sont détaillées dans les parties suivantes à travers leur mise en oeuvre pour la sélection des réseaux routier et hydrographique.

Les hypothèses faites sur les données pour que cette approche soit applicable sont issues des données de test (celles utilisées dans les premiers, voir plus haut) mais ont été étendues pour des besoins de généralité. L'hypothèse forte est l'utilisation de réseaux entièrement linéaires (tous les éléments du réseau sont linéaires et il n'y a pas de partie surfacique comme un lac dans un réseau hydrographique) et topologiquement corrects. Aucun attribut n'est indispensable sur les éléments de réseau même si l'utilisation des attributs présents dans les données peut améliorer la qualité des résultats comme dans le cas des cours d'eau dont le nom peut être utilisé pour former les "cours d'eau continus" (voir §3.1.1).

3. Enrichissement de données par analyse spatiale

Dans le cadre de la méthode de sélection que nous proposons, l'enrichissement de données par analyse spatiale a deux objectifs principaux : faciliter et guider le processus de sélection standard d'une part et expliciter des structures caractéristiques du réseau d'autre part qui pourront être maintenues par notre méthode générale (par l'étape de typification). Ces objectifs sont illustrés dans les paragraphes suivants par des méthodes de détection de structures dans les réseaux routiers et hydrographiques puis des méthodes de caractérisation de ces structures.

3.1. Détection de structures dans le réseau

L'étape préalable à tout enrichissement est l'identification théorique des structures spatiales importantes pour notre application (ici la généralisation de base de données) et de leurs relations. Cette identification a été réalisée par les premiers tests effectués et par analyse de la littérature. La première étape est ensuite de détecter ces structures par des méthodes d'analyse spatiale. La première partie de cette section présente des structures détectées par des méthodes directement adaptées de la littérature. Les deux parties suivantes détaillent des méthodes que nous avons développées pour d'autres structures caractéristiques des réseaux routiers et hydrographiques. La dernière partie propose une évaluation de ces méthodes d'enrichissement.

3.1.1. Adaptation de méthodes connues

Dans le cas des réseaux, l'enrichissement clé pour servir de base à la sélection est le recours au principe de bonne continuité du Gestalt (Thomson et Richardson, 1999). Les psychologues du Gestalt ont étudié les différents mécanismes cérébraux du regroupement perceptuel (Wertheimer, 1923). Le principe de bonne continuité permet de créer des groupes perceptuels appelés "strokes" (Thomson et Richardson, 1999) qui sont des ensembles de tronçons qui semblent se suivre de manière continue comme si on les avaient dessinés d'un seul trait. Ces strokes ont une vraie signification géographique car les réseaux sont des phénomènes assez continus (Figure 3) : en effet, un géographe parle de *routes* ou de *cours d'eau* et non de tronçons, qui sont les objets présents initialement dans la base de données; ces strokes font office de *routes* ou de *cours d'eau*. Le problème principal de la création des strokes est la détermination de la continuité à un nœud du réseau. Au-delà de la continuité purement géométrique calculée par estimation de la courbure au nœud, nous avons spécialisé des méthodes de détermination de la continuité pour les réseaux routier et hydrographique. Dans le cas des réseaux hydrographiques, nous parlons de "cours d'eau continus" car le principe de continuité est ici calculé dans le sens d'écoulement de l'eau (information nécessaire) et contraint par des continuités attributaires en fonction des attributs disponibles comme le nom du cours d'eau mais également quand le réseau contient les sources des cours d'eau, par la distance à la source (critère utilisé dans la détermination des cours d'eau par les géographes). Dans le cas des routes, la méthode est enrichie pour gérer sans interruption la continuité à la traversée des ronds-points qui ont été préalablement détectés par la méthode proposée dans (Sheeren, 2002) (Figure 5a). Dans notre méthode, une partie de la sélection est ensuite appliquée sur les strokes plutôt que sur les tronçons.

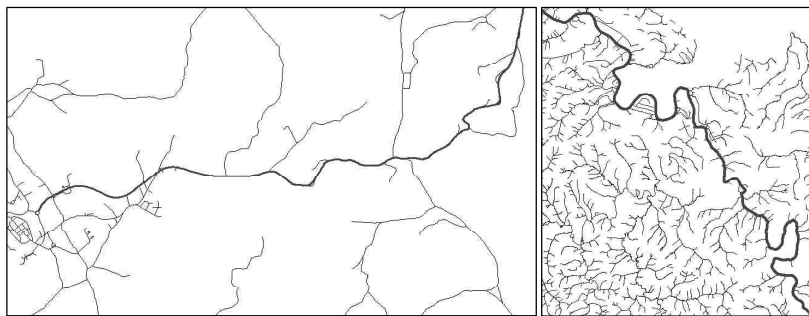


Figure 3. A gauche, un stroke de réseau routier et à droite un stroke de réseau hydrographique ou "cours d'eau continu".

L'identification des impasses dans le réseau routier est également un enrichissement important. De nombreuses spécifications de sélection distinguent en effet les impasses des autres routes. Nous avons repris la méthode de (Duchêne, 2005) qui considère le réseau routier comme un graphe et s'intéresse aux faces de ce

graphe. Les tronçons se trouvant à l'intérieur d'une face sont considérés comme des impasses. Cette méthode permet de reconnaître des impasses complexes (Figure 4). Un autre enrichissement non directement lié au réseau routier sert de base au processus de sélection : la délimitation de l'étendue des villes (Figure 4) permet d'utiliser des processus de sélection standard différents en ville et en zone rurale puisque des travaux ont montré que c'était plus efficace. Nous avons utilisé parmi les différentes méthodes existantes celle de (Boffet, 2002) : des buffers de 50 m (seuil déterminé empiriquement) sont appliqués aux bâtiments (les polygones de la Figure 4) puis les buffers qui s'intersectent sont agrégés. Les petits groupes ainsi formés sont éliminés et le contour des grands groupes est simplifié pour obtenir les limites visibles sur la Figure 4.



Figure 4. A gauche, création des objets villes par la méthode de (Boffet, 2002) et à droite des configurations d'impasses complexes détectées (Duchêne, 2005).

3.1.2. Détection de structures routières par des algorithmes ad hoc

En plus d'adapter des méthodes de la littérature, nous avons développé des méthodes ad hoc lorsque c'était nécessaire. Premièrement, l'identification et la classification des carrefours simples (croisements de tronçons routiers) et des carrefours complexes (configuration particulières de carrefours simples) permet à la fois de guider le processus de sélection standard et d'isoler des structures caractéristiques qui devront être exclues du processus de sélection standard pour être typifiées (Figure 2).

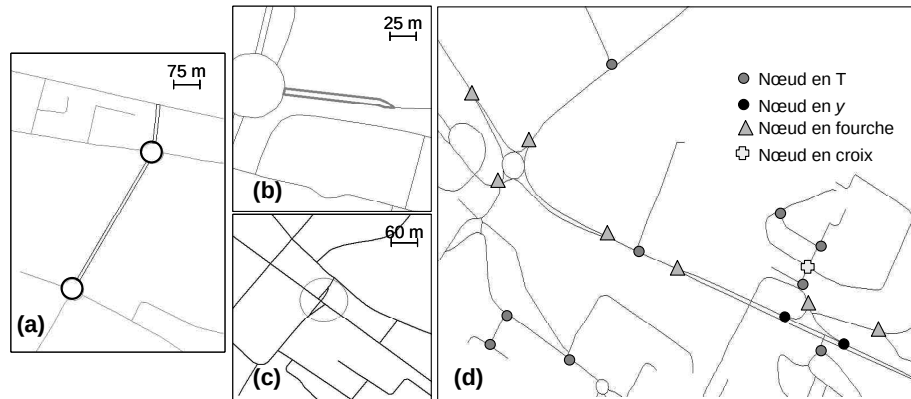


Figure 5. Classification automatique de carrefours: (a) ronds-points et voies chaussées séparées (b) patte d'oie (c) carrefour décalé (d) carrefours simples.

Notre classification des carrefours simples s'inspire de (Grosso, 2004) et distingue les carrefours en fourche, en y, en T, en croix ou en étoile (Figure 5d). Les carrefours complexes pour lesquels nous avons développé une méthode de détection sont les pattes d'oies, les voies à chaussées séparées et les carrefours décalés (Figure 5). Le détail des algorithmes permettant ces enrichissements est présenté dans (Touya, 2007a).

Nous présentons maintenant des structures autres que les carrefours. Les réseaux routiers sont parfois utilisés pour des applications de navigation pour lesquelles la détection des aires d'autoroute est nécessaire. Elles se caractérisent par un ensemble de tronçons de routes connectés entre eux dont l'un permet de sortir de l'autoroute et un autre d'y rentrer. Nous les détectons en deux temps : nous repérons d'abord les entrées et sorties d'autoroutes à partir des nœuds en y connectés à une autoroute et à partir de l'orientation du carrefour par rapport à cette autoroute (Figure 6a) ; nous identifions ensuite les petites faces du graphe routier situées du bon côté de l'autoroute entre une sortie et une entrée. Nous prenons alors l'enveloppe convexe de l'ensemble des tronçons de ces faces pour obtenir l'étendue de l'aire d'autoroute (Figure 6b). Quand deux aires d'un côté et de l'autre d'une autoroute sont reliées par un tronçon (un pont ou un tunnel), comme dans l'exemple de la Figure 6b, elles sont agrégées.

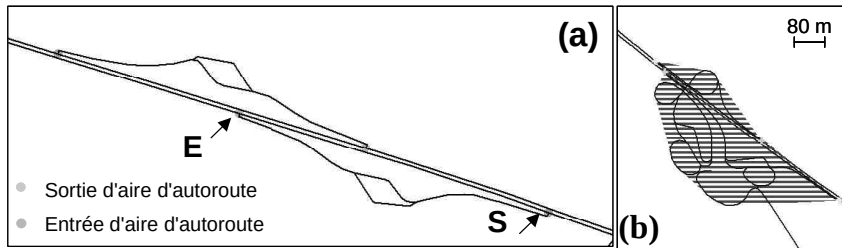


Figure 6. (a) *identification automatique des entrées et sorties d'aires d'autoroute.* (b) *détection de la structure formée de l'ensemble de tronçons routiers.*

Les échangeurs autoroutiers sont une autre structure caractéristique des réseaux routiers. D'après (Grosso, 2004, Heinzle et Anders, 2007), ils se caractérisent par une forte densité de nœuds en fourche et en y (Figure 7). Nous exploitons cette règle en formant des groupes de nœuds par agglomération de proche en proche, en utilisant la distance curviligne par le réseau. Ces groupes sont ensuite discriminés suivant la présence de bâtiments à l'intérieur de l'enveloppe du groupe (il n'y a généralement pas de bâtiment au milieu d'un échangeur). Toutes les routes incluses dans l'enveloppe convexe du groupe de nœuds, ainsi que celles connectées topologiquement au groupe composent l'échangeur routier.

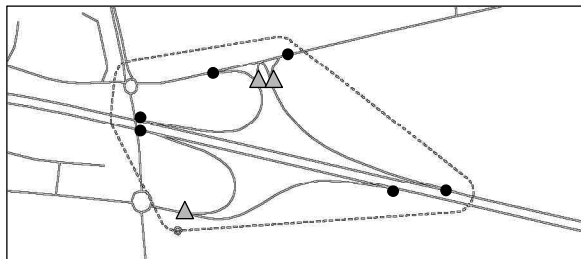


Figure 7. *Un échangeur détecté automatiquement par agglomération de nœuds en y (points) et en fourche (triangles), matérialisé par l'enveloppe de ses nœuds.*

3.1.3. Détection de structures hydrographiques

Dans un réseau hydrographique linéaire, les différents bras d'une rivière sont souvent représentés dans la base de données, marquant la présence d'une ou plusieurs îles sur la rivière. Nous considérons ces îles comme des structures caractéristiques des réseaux hydrographiques que nous détectons afin de les maintenir ou de les simplifier de façon pertinente en fonction des spécifications. Les îles simples sont les petites faces du graphe hydrographique. Les îles complexes sont détectées par agglomération d'îles simples adjacentes (Figure 8). Le contour est alors déterminé par voisinage topologique (les tronçons de cours d'eau extérieurs).

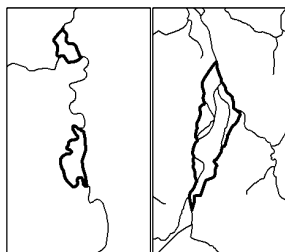


Figure 8. Des îles simples à gauche et une île complexe composée d'îles adjacentes à droite. Le contour est créé automatiquement.

Nous considérons également que les zones d'irrigation sont des structures caractéristiques dans un réseau hydrographique, qui influent sur la sélection et nécessitent un traitement particulier car elles n'appartiennent pas vraiment au réseau naturel, traité par la sélection standard. Il s'agit, dans des zones plates, de canaux artificiels et souvent tracés de manière rectiligne qui permettent l'irrigation des champs. De plus, les tronçons étant plats, les sens d'écoulement assignés dans ces zones sont hautement incertains arbitraires. Nous les détectons par classification ascendante hiérarchique : dans les zones plates, les sources et les pertes (extrémités du réseau) sont regroupées selon la distance, la classification étant stoppée au-delà d'un seuil de distance. Seuls les groupes assez grands sont conservés. L'enveloppe convexe de l'ensemble des tronçons inclus dans ces groupes forme la délimitation des zones d'irrigation. Les groupes sont ensuite discriminés suivant leur forme et suivant la densité en petits tronçons pour ne conserver ceux qui sont assez compacts et très denses (Figure 9). L'interprétation des différents seuils est complexe ce qui rend difficile l'obtention automatique des toutes les zones sans conserver quelques artefacts (des zones trouvées qui ne sont pas des zones d'irrigations).

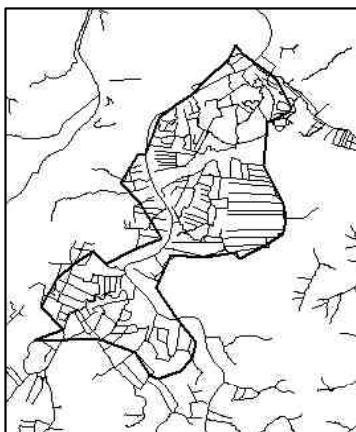


Figure 9. Une zone d'irrigation détectée automatiquement par notre méthode.

Les méandres sont des éléments caractéristiques de la morphologie des rivières. Nous les avons détectés pour guider la sélection mais ils pourraient servir à des analyses géomorphologiques des réseaux hydrographiques. Nous les avons détectés en segmentant les "cours d'eau continus" (voir §3.1.1) en parties homogènes en sinuosité selon une adaptation de la méthode de (Plazanet, 1995) destinée plutôt aux routes de montagne. Les virages des parties sinueuses sont ensuite qualifiés, par des mesures classiques comme le nombre de points d'inflexion ou le rapport base/longueur, pour trouver les méandres de petite ou grande ampleur (Figure 10).

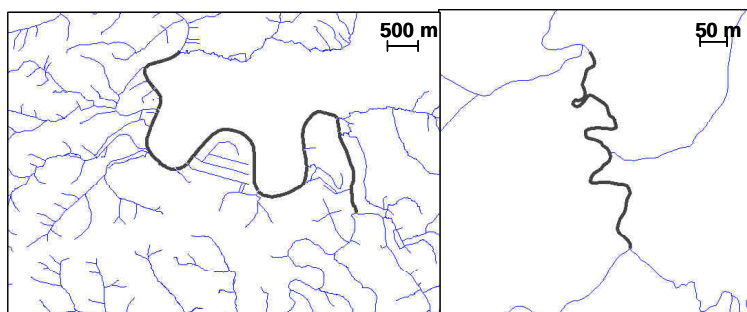


Figure 10. *A gauche, des grands méandres détectés automatiquement et à droite des petits méandres.*

3.1.4. *Evaluation des méthodes d'enrichissement*

Les méthodes de détection automatique des structures (dont les structures caractéristiques) ont été évaluées par des tests et les résultats sont jugés satisfaisant. Les tests ont été effectués sur tout un département (Pyrénées-Atlantiques) issu de la BD Topo® de l'IGN. Puis, les enrichissements ont été appliqués sur d'autres zones de cette même base de données pour valider la démarche. Les différentes images présentées dans cette partie 3.1 sont issues de ces résultats. Pratiquement toutes les structures sont détectées. A titre d'exemple, après validation visuelle, la méthode de détection trouve bien les 269 ronds-points de la zone et pas un de plus. Mais une aire de repos sur 13 est sous-détectée car l'entrée est située exactement au niveau d'un pont (ce qui donne un nœud de degré 5 qui n'est pas choisi comme entrée). L'algorithme est donc à peaufiner quelque peu pour améliorer les résultats. Il y a aussi parfois un peu de sur-détection. Ainsi, pour les zones d'irrigations 2 des 20 zones détectées n'en sont pas après vérifications visuelle. De plus, en travaillant sur un département entier (100 000 tronçons routiers et 50 000 tronçons hydrographiques), nous pouvons assurer des temps de calcul faibles (moins d'une minute pour détecter les aires de repos par exemple).

3.2. Caractérisation des structures détectées

Une fois les structures intéressantes détectées, il convient de les caractériser pour les exploiter au mieux dans le processus de sélection, que ce soit pour guider la sélection standard ou pour faciliter la typification. La caractérisation des structures revient à mesurer, qualifier les différents composants de la structure et les relations entre eux et avec les objets extérieurs de la structure. Par exemple, une fois détectée une structure telle qu'une zone d'irrigation, sa caractérisation consiste à identifier ses composants, les mesurer pour déterminer ceux qui sont naturels (des tronçons de cours d'eau) et ceux qui sont artificiels (des canaux d'irrigation). La première partie présente des méthodes de caractérisation de structures pour illustrer l'intérêt de cette caractérisation. La deuxième partie présente l'intérêt d'enrichir également le schéma de données pour gérer la détection et la caractérisation des structures.

3.2.1. La caractérisation de structures de réseaux

Les carrefours complexes (Figure 5) sont caractérisés en discriminant par topologie les tronçons routiers qui sont intérieurs au carrefour (ceux en rond dans un rond-point) et les tronçons qui permettent d'accéder au carrefour, mais aussi en mesurant la forme du carrefour. Cette caractérisation a permis ainsi de développer une méthode de construction des strokes qui prend en compte ronds-points et pattes d'oies : les tronçons intérieurs ne sont pas considérés et seuls les tronçons d'accès au carrefour complexe sont pris en compte dans le calcul de continuité (un tronçon artificiel est ajouté pour traverser le carrefour). Cette caractérisation permet aussi de guider plus précisément le processus de sélection du réseau routier en zone urbaine présenté en partie 4.1. Enfin, cette caractérisation est fondamentale pour la typification de ces structures comme on peut le voir dans la partie consacrée à la phase de typification (§4.3).

De même, la caractérisation précise des carrefours en T est utile pour déterminer facilement les strokes routiers qui sont importants car ils desservent beaucoup de routes secondaires (Figure 11). Les tronçons formant le haut du T font partie d'une route plus importante que celui formant la barre du T. Si un stroke passe par beaucoup de carrefours en T en empruntant les tronçons importants, comme dans la Figure 11, il est considéré comme important.



Figure 11. Le stroke sélectionné (en tirets) est important car il dessert beaucoup de routes secondaires par des carrefours en T.

Dans le cas des zones d'irrigation des réseaux hydrographiques, les tronçons inclus dans la zone d'irrigation sont discriminés en tronçons naturels (longs et sinueux) et en tronçons artificiels (courts et droits) car les réseaux d'irrigation s'appuient sur le réseau naturel et les deux types de tronçons coexistent dans une zone d'irrigation. Cette distinction automatique (Figure 12) permet au processus de sélection standard des cours d'eau de ne prendre en compte que les tronçons naturels pour ne pas voir de résultats biaisés. Le réseau d'irrigation proprement dit doit être traité par un processus de typification spécifique.

La caractérisation précise des impasses complexes (Figure 4) sous forme d'arbre permet guider leur sélection en ne gardant que les branches les plus longues de l'arbre. Cette caractérisation se fait en déterminant la racine (le nœud connectant la structure au reste du réseau) et les extrémités du groupe puis en réalisant des mesures de plus court chemin entre la racine et les extrémités. Une caractérisation identique permet de typifier des structures analogues dans d'autres types de réseaux comme les deltas dans un réseau hydrographique (non détectés ici car non présents dans nos données test) ou les gares de triage dans un réseau ferré.



Figure 12. Discrimination des tronçons naturels (épais) et artificiels (fins) dans une zone d'irrigation (contour noir).

3.2.2. Un schéma de travail enrichi

(Ruas, 2004, p.42) explique qu'en plus du schéma de données final des données (celui de la base de données généralisée), la généralisation nécessite d'utiliser temporairement un schéma de données enrichi appelé *schéma de travail*. Dans notre cas, il est intéressant d'enrichir le schéma de travail avec les structures détectées et leur caractérisation. C'est ce schéma de données de travail enrichi qui permet de

faciliter et de guider la sélection ainsi que la phase de typification. La Figure 13 montre un rond-point connecté à une patte d'oie dans un réseau routier ainsi que le schéma de données enrichi pour stocker les deux structures et leur caractérisation décrite dans la partie précédente. Le stockage de cet information au début du processus permet de l'utiliser facilement et rapidement dans les deux étapes suivantes.

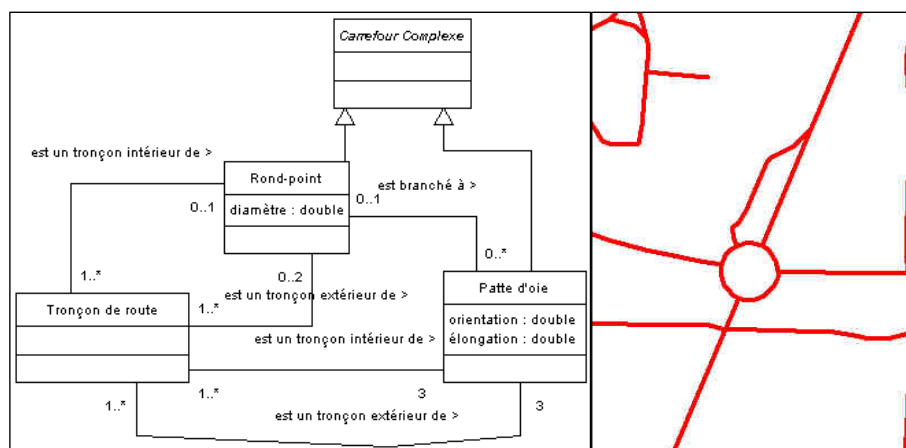


Figure 13. Schéma de données enrichi explicitant structures et relations implicites concernant les situations telles que celle de l'image de droite: un rond-point et une patte d'oie.

4. Processus de sélection standard

Cette partie présente de manière succincte les processus de sélection standard s'appuyant sur les enrichissements précédents, développés pour les réseaux routiers et hydrographiques. Ces processus sont plus détaillés dans (Touya, 2007a) pour les routes et (Touya, 2007b) pour les cours d'eau.

4.1. La sélection des routes

Comme préconisé par (Ruas, 1999, p.163) du fait de différences fondamentales entre les réseaux routiers urbains et ruraux (forme, densité, fonction), le processus de sélection standard que nous proposons pour les routes est séparé en trois phases : la sélection en zone rurale, la sélection en zone urbaine et la vérification de la continuité à l'interface des deux zones.

Notre processus de sélection en zone rurale reprend et complète celui de (Richardson et Thomson, 1996). La sélection est ainsi basée sur l'évaluation de la fréquentation des routes par calcul de plus courts chemins entre des points

d'attraction (aéroport, école, zone d'habitat résidentiel). Mais notre sélection s'appuie également sur les strokes qui sont qualifiés par leur longueur et leur nombre d'intersection (voir §3.2.2) et sur les impasses. Sont sélectionnées les routes à forte fréquentation ou appartenant à un stroke important ainsi que les impasses assez longues (voir §3.2.2).

La sélection en zone urbaine s'appuie sur une approche complètement différente qui poursuit les travaux de (Ruas, 1999, p.163). L'idée est d'agréger les faces du graphe routier et d'éliminer les rues séparant deux faces agrégées. L'agrégation se base sur la fréquentation et la longueur du stroke de la rue séparatrice mais aussi sur la taille et la forme de la face agrégée.

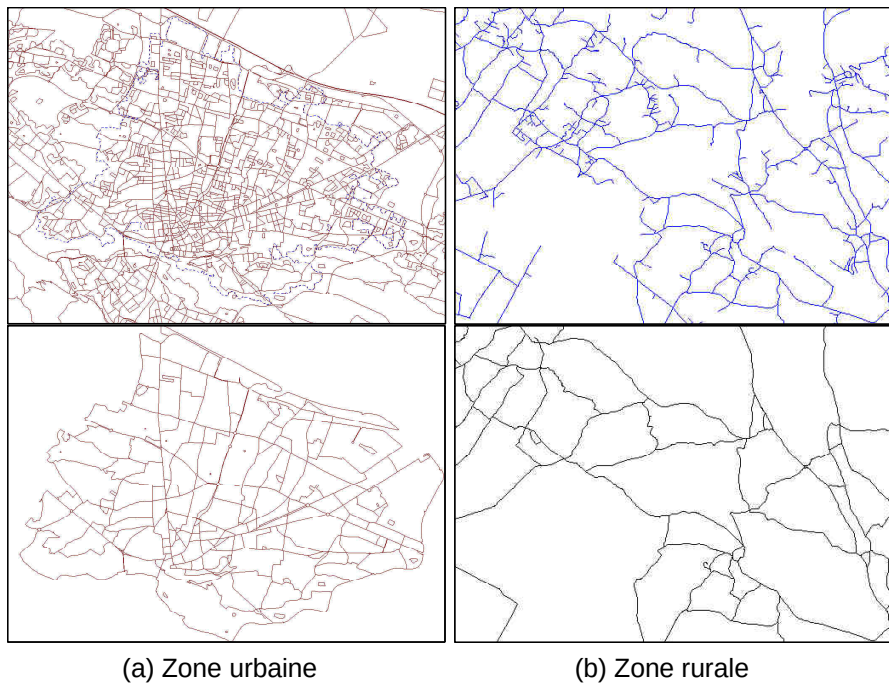


Figure 14. (a) Une zone urbaine et sa sélection en dessous. (b) Une zone rurale et sa sélection (Touya, 2007a).

Enfin, la continuité du réseau à l'interface des deux zones est assurée par un phase de vérification qui s'appuie sur les strokes. Parmi les strokes qui traversent deux zones, s'il y en a dont seulement une partie des tronçons a été sélectionnée par un des deux processus, on sélectionne après coup le reste du stroke pour assurer la continuité.

Cette méthode de sélection ainsi que les outils d'enrichissement ont été implémentés sur le SIG Clarity™ de la société 1Spatial. Des tests ont été effectués sur le réseau routier de la BD Topo® de l'IGN, de résolution métrique, avec des paramètres correspondant à différentes spécifications en sortie (pour des résolutions de bases de données différentes). La Figure 14 montre le résultat en zone rurale et urbaine sur un extrait pour des spécifications correspondant à une autre base de données de l'IGN, la BD Carto® de résolution 10m. Le processus a donc été paramétré pour ces spécifications permettant, pour évaluation, la comparaison avec cette base. On obtient quasiment les mêmes tronçons en zone rurale (97 % d'arcs identiques dans la zone test) et une sélection équivalente en zone urbaine (les mêmes tronçons importants et peu d'écart, ici 2%, en longueur totale de tronçons). Nous estimons ces résultats globalement satisfaisants.

4.2. La sélection des cours d'eau

Comme dans (Thomson et Brooks, 2000), la sélection des cours d'eau consiste à réaliser une classification de Horton (Horton, 1945) sur les cours d'eau continus (Figure 15). Dans le réseau schématisé de la Figure 15, le critère principal de continuité à un nœud est de continuer celui qui est le plus long de sa source au nœud. Un ordre d'importance est attribué à chaque cours d'eau continu en fonction du nombre d'affluents. Nous sélectionnons ensuite les cours d'eau continus dont l'ordre d'importance est supérieur à un seuil. De plus cette sélection est contrôlée par les informations issues de l'enrichissement. Ainsi, les îles, les deltas, les zones d'irrigation ou les méandres contraignent la sélection. Un exemple des contraintes que nous avons ajoutées est le suivant : si on sélectionne une partie d'un méandre, on doit sélectionner l'ensemble du méandre.

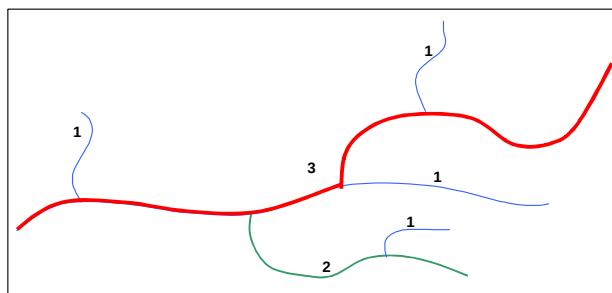


Figure 15. Classification de Horton de cours d'eau continus (créés manuellement sur ce schéma) sur laquelle s'appuie la sélection des cours d'eau.

Comme pour le réseau routier, nous avons implémenté notre méthode de sélection de réseau hydrographique sur le SIG Clarity™ et l'avons testé sur le réseau de la BD Topo® pour différentes résolutions de sortie. La Figure 16 montre un extrait d'un des tests de sélection paramétré pour des spécifications correspondant à une base de résolution 25m. La méthode a été évaluée en réalisant une comparaison

avec une autre base de données de l'IGN, la BD Carto® de résolution 10m. Dans le test utilisant comme paramètres les spécifications de cette base, on obtient un résultat équivalent (92 % de longueur commune dont les mêmes cours d'eau principaux) mais surtout plus cohérent avec ses spécifications par endroits. La notion de sélection équivalente signifie que les éléments les plus importants sont toujours bien sélectionnés et que la même quantité de réseau secondaire a été sélectionnée même si ce ne sont pas exactement les mêmes éléments (Touya *et al*, 2006). La sélection pourrait être améliorée en étant intégrée à un processus plus complet traitant également les lacs et les cours d'eau surfaciques.

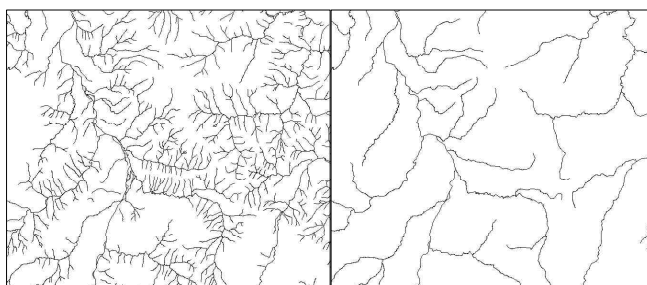


Figure 16. Résultats de la sélection sur des données BD Topo® (résolution 1m) vers des données adaptées à une base de résolution 25m (Touya, 2007b).

4.3. La phase de typification

Pour les structures caractéristiques du réseau (§3.2), nous avons indiqué que les processus de sélection généraux ne sont pas adaptés pour maintenir correctement la structure. Des méthodes de typification ad hoc sont nécessaires. Typifier consiste à "porter dans une présentation, qui en préserve la distribution reconnue, des objets de même nature qui se trouvent groupés localement par quelque processus géographique identifiable" (Hangouët, 1998, p. 227). Trois exemples sont présentés dans cette partie : les îles complexes, les échangeurs routiers et les carrefours complexes.

Pour permettre plusieurs niveaux d'abstraction dans la généralisation des îles complexes, il est intéressant d'avoir recours à une agrégation hiérarchique : par exemple, les plus petites îles sont agrégées à une île voisine selon leur taille (Figure 17).

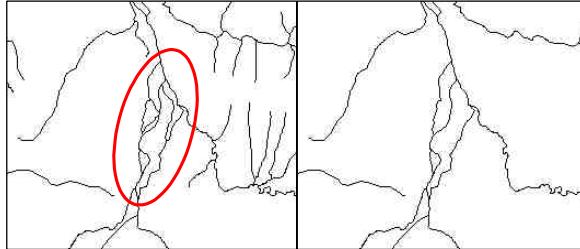


Figure 17. Une île complexe typifiée : les bras principaux sont conservés et le phénomène d'île complexe est préservé.

Pour les échangeurs routiers, nous avons adapté l'algorithme de (Mackaness et Mackechnie, 1999) qui agrège les nœuds constitutifs de l'échangeur de manière hiérarchique. Cela permet ensuite plusieurs niveaux d'élimination des nœuds dans l'échangeur (Culson, 2006) et permet de simplifier l'échangeur (Figure 18). Certains segments recréés durant la typification nécessitent une correction géométrique comme dans le cas de l'ordre 1 sur la Figure 18.

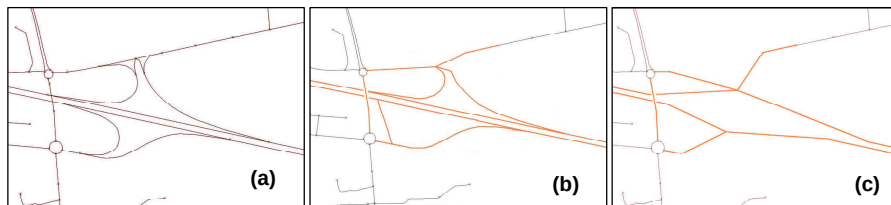


Figure 18. un échangeur routier (a) et ses formes typifiées à différents degrés (b) et (c). Des corrections géométriques sont nécessaires pour corriger les déformations.

Bien caractériser une structure avec ses composants et ses relations (voir §3.2.2) permet dans de nombreux cas de faciliter sa typification. Par exemple, la modélisation enrichie des ronds-points et pattes d'oie de la Figure 13 permet, une fois instanciée, de typifier les ronds-points en carrefour en croix (selon le diamètre) et les pattes d'oies en carrefour en T très facilement (en utilisant les tronçons intérieurs et extérieurs). La Figure 19 montre des résultats de ces typifications simples.

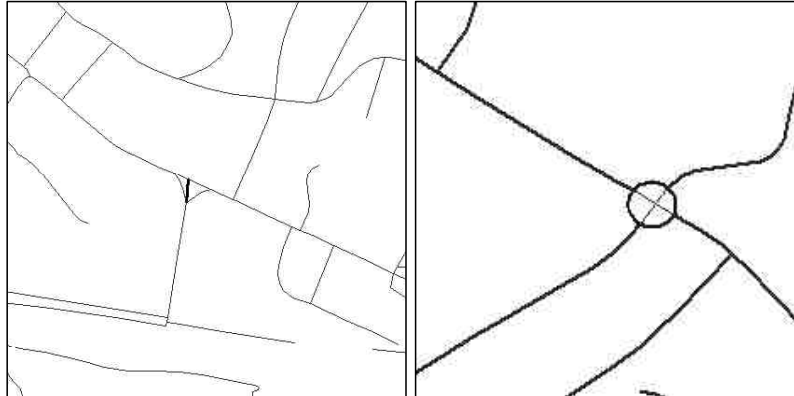


Figure 19. *A gauche, une patte d'oie typifiée en carrefour en T (le trait en gras remplace les deux pattes). A droite, un rond-point typifié en carrefour en croix (les traits fins remplacent la structure).*

5. Autres applications possibles des méthodes d'enrichissement proposées

Les enrichissements présentés dans cet article ont été conçus dans le contexte de la sélection de réseaux. Mais la plupart d'entre eux peuvent être utiles dans le cadre d'autres applications de géomatique. La partie 5.1 identifie des enrichissements développés pour la sélection pouvant servir à des analyses spatiales géographiques. La partie 5.2 décrit comment d'autres enrichissements peuvent s'insérer dans d'autres applications de géomatique que la généralisation de base de données. La dernière partie traite des bases de données multi-représentations.

5.1. Application à des analyses géographiques

(Paget *et al*, 2008) montre que l'analyse des structures du réseau hydrographique comme les méandres ou deltas, ou l'analyse des caractéristiques des cours d'eau continus, peuvent donner des indices sur les structures géomorphologiques sous-jacentes. La Figure 20 montre la classification des structures géomorphologiques de réseaux hydrographiques de (Howard, 1967) qui pourrait être construite automatiquement à l'aide de structures enrichies du réseau. La caractérisation des strokes avec par exemple l'orientation générale pourrait permettre de discriminer les réseaux de forme parallèle (même orientation) ou de forme radiale (orientation dirigée vers un centre commun).

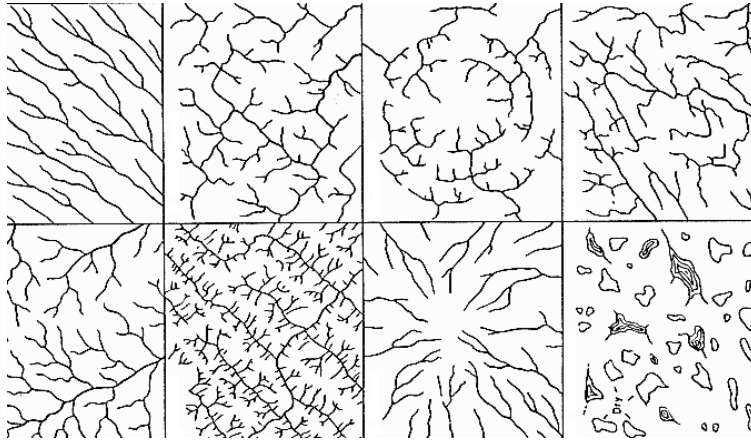


Figure 20. Différentes structures géomorphologiques de réseaux hydrographiques selon la classification de Howard (Howard, 1967).

De même, la détection d'impasses complexes ou de structures maillées dans un réseau routier peuvent servir de support à l'analyse géographique du tissu urbain. Ainsi, certaines des méthodes d'analyse de réseaux présentées dans cet article seront utilisées dans le projet ANR en cours GeOpenSim (GeOpenSim, 2008; Perret *et al*, 2009) qui vise à construire une plate-forme open source dédiée à l'analyse des tissus urbains. Par exemple, la Figure 21 montre comment la détection et la caractérisation d'un groupe complexe d'impasses particulier peut être caractéristique d'une zone industrielle : le réseau routier donne des informations sur les formes urbaines.

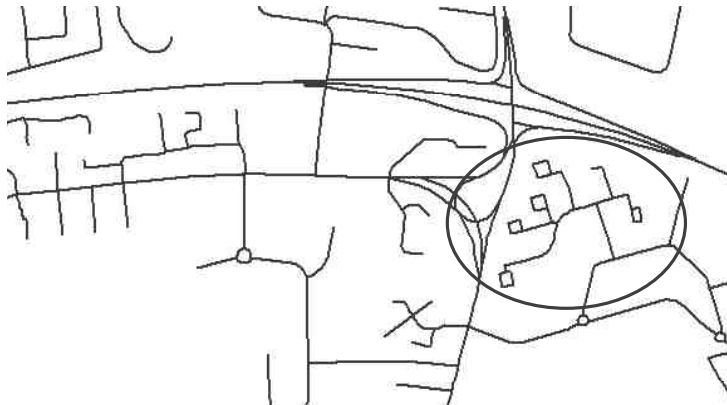


Figure 21. Le groupe d'impasses complexe entouré permet de détecter et caractériser une zone industrielle.

5.2. Application à d'autres traitements de géomatique

Les enrichissements que nous avons présentés dans cet article peuvent faciliter et guider d'autres processus de géomatique que la généralisation de base de données. Ils sont tout d'abord très utiles dans le cadre de la généralisation à but cartographique (le but est de produire une carte à une échelle donnée). En effet, la généralisation de base de données est souvent vu comme une étape de la généralisation cartographique et dans ce cas, les enrichissements sont doublement utiles. Ils sont d'abord utiles pour sélectionner les bonnes informations. Mais la détection de structures routières comme les ronds-points, pattes d'oies et voies à chaussées séparées est aussi un préalable indispensable pour guider la généralisation cartographique par des processus comme ceux de (Harrie, 1999, Bader et Barrault, 2001) si l'on veut éviter des déformations trop importantes. En effet, ces processus optimisent la déformation et le déplacement des objets symbolisés et détecter les ronds-points permet de ne pas les déformer.

Par ailleurs, ces mêmes structures sont utilisées comme entrée de processus d'appariement de données géographiques comme celui de (Mustière et Devogele, 2008) (Figure 22). Le même procédé peut être utilisé pour l'appariement de réseaux hydrographiques avec les îles simples et complexes.

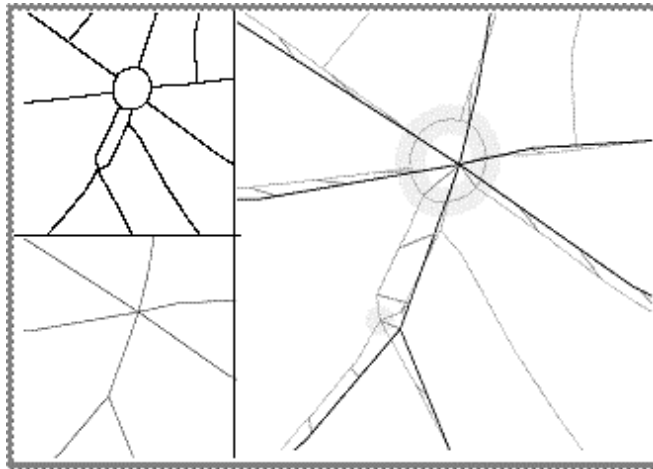


Figure 22. A gauche, deux bases de données au niveau d'un carrefour complexe avec rond-point et patte d'oie. Leur détection permet l'appariement correct de leurs nœuds et arcs sur l'image de droite (Mustière et Devogele, 2008).

5.3. Applications à la représentation multiple

Enrichir une base de données par des structures qui représentent des concepts géographiques implicites à la résolution de la base de données est un bon point de

départ à la constitution d'une base de données multi-représentations. Les bases de données multi-représentations permettent de faire coexister plusieurs points de vue et résolutions spatiales (Vangenot *et al*, 2002). Pour une même entité géographique du monde réel, il peut exister plusieurs représentations en fonction des points de vue ou des résolutions. Par exemple, une ville est représentée par un réseau de rues et des bâtiments à grande échelle, par un polygone représentant son étendue à petite échelle ou par un nœud dans selon un point de vue navigation routière. La constitution d'une base de données multi-représentations est intimement liée aux applications décrites dans la partie précédente car les méthodes courantes pour constituer automatiquement de telles bases reposent sur la généralisation et l'appariement. Il apparaît donc évident que les enrichissements décrits dans cet article sont par conséquent utiles à la constitution de bases de données multi-représentations.

Les aires de repos sont un bon exemple d'utilisation de nos enrichissements dans le cadre d'une base de donnée multi-représentations : posséder un algorithme de détection de la structure permet d'intégrer le concept dans la base de données pour de l'instancier; sa caractérisation (non présentée dans l'article : discrimination des nœuds et des tronçons d'entrée et sortie de l'aire) peut permettre une représentation simplifiée à une résolution plus grande; ou la non représentation de l'objet à certaines résolutions où cela n'a plus de sens; enfin, si un point de vue navigation routière est utilisé, l'objet aire de repos peut être représenté facilement comme une information attachée à des nœuds des autoroutes (au niveau des entrées par exemple).

6. Conclusion et perspectives

En conclusion, cet article présente une méthode générale de sélection des réseaux basé sur l'enrichissement de données par analyse spatiale. Les contribution de ce travail sont de plusieurs ordres : d'abord la sélection et l'adaptation de méthodes d'enrichissement issues de la littérature; le développement de méthodes d'enrichissement complémentaires; le développement de nouvelles méthodes de typification; l'adaptation de méthodes de sélection de la littérature en les guidant à l'aide des méthodes d'enrichissement; enfin, la conception de la méthode générale de sélection des réseaux basée sur l'enrichissement. La méthode a été mise en oeuvre pour la sélection des réseaux routiers et hydrographiques et testée avec succès sur des données topographiques de l'IGN. Les enrichissements développés et adaptés peuvent également être utiles pour d'autres applications que la généralisation comme des analyses géographiques ou des dérivations spécifiques.

Les perspectives sont de trois ordres. Il serait d'abord intéressant de mettre en oeuvre cette méthode sur d'autres réseaux géographiques comme les réseaux ferrés ou électriques. En effet, ces réseaux ont aussi besoin d'être généralisés et cela permettrait de bien tester la généricité de la méthode en étendant le champ d'application. Par ailleurs, il serait évidemment intéressant de développer des méthodes d'enrichissement pour d'autres structures des réseaux hydrographiques et

routiers comme les deltas ou les boulevards périphériques (Heinzle et Anders, 2008). Cela permettrait de contrôler encore mieux les processus de sélection. Enfin, il serait intéressant d'approfondir la modélisation des objectifs d'une généralisation et notamment la traduction des volontés d'un utilisateur en paramètres de ces processus de sélection. L'utilisation de méthodes d'aide au paramétrage basées sur des dialogues à partir d'échantillons comme (Hubert, 2003, 2008; Christophe, 2008) permettrait à un utilisateur standard de spécifier ses paramètres de sélection. Le recours aux échantillons est d'ailleurs une solution envisagée de manière générale pour le paramétrage des processus de généralisation complexes comme dans le cas de modèle de généralisation cartographique (Duchêne et Gaffuri, 2008).

Bibliographie

- Bader, M., Barrault, M., « Cartographic Displacement in Generalization: Introducing Elastic Beams ». *In proceedings of 4th workshop on progress in automated map generalisation*, Beijing, 2001.
- Boffet, A., « Analyse multi-niveaux des espaces urbains ». *Revue Internationale de Géomatique*, Vol. 12, n° 2, 2002, pp. 215-260.
- Christophe, S., « Creative cartography based on dialogue ». *In the proceedings of the Conference AutoCarto 2008*, 8-10 September
- Culson, N., « Détection et généralisation des échangeurs routiers dans une base de données cartographique vectorielle ». Mémoire de DESS, Université Paris 6. Laboratoire COGIT, IGN. 2006.
- Duchêne, C., « Enrichissement de données urbaines par calculs d'analyse spatiale ». Rapport interne du laboratoire COGIT, IGN. 2005.
- Duchêne, C., Gaffuri, J., « Combining Three Multi-Agent Based Generalisation Models: AGENT, CartACom and GAEL ». *In proceedings of the 13th International Symposium on Spatial Data Handling*, (SDH'08), Montpellier, France. 2008.
- GeOpenSim, « Les objectifs du projet GeOpenSim ». <http://geopensim.ign.fr>
- Grosso, E., « Etude des carrefours d'un réseau routier – Première approche : modélisation et implémentation d'outils de détection des carrefours (plate-forme Oxygène) ». Mémoire de DESS, Université Paris 1. Laboratoire COGIT, Institut Géographique National. 2004.
- Hangouët, J.-F., « Approche et méthodes pour l'automatisation de la généralisation cartographique; application en bord de ville ». Thèse de doctorat, Université de Marne-la-Vallée, 1998.
- Harrie, L.E., « The Constraint Method for Solving Spatial Conflicts in Cartographic Generalization ». *Cartography and Geographic Information Science (CaGIS)*, Vol. 26, n°1, 1999, pp. 55-69.

- Heinzle, F., Anders, K.-H., « Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks ». Mackaness W., Ruas A., Sarjakoski T. (eds) : *The Generalisation of Geographic Information : Models and Applications*. Elsevier (2007).
- Horton, R.A., « Erosional development of Streams and their Drainage Basins:Hydrophysical approach to Quantitative Morphology ». *Geo. Soc. America Bull*, Vol. 56 (1945) 275-370
- Howard, A.D., Drainage Analysis in Geologic Interpretation: a summation. *American Association of Petroleum Geologists Bulletin*, Vol. 51 (11), 1967, pp. 2246-2259.
- Hubert, F., « Modèle de traduction des besoins d'un utilisateur pour la dérivation de données géographiques et leur symbolisation par le Web ». Thèse de doctorat soutenue à l'Université de Caen. 2003.
- Hubert, F., « CartAble, système d'aide au paramétrage de traitements géographiques complexes ». *Revue Internationale de Géomatique*, Vol. 18, n°1, 2008.
- Mackaness, W., Edwards, G., « The Importance of Modelling Pattern and Structure in Automated Map Generalisation ». In *Proceedings of Joint Workshop on Multi-scale Representations of Spatial Data*, Ottawa, Canada (2002).
- Mackaness, W., Mackechnie, G.A., « Automating the Detection and Simplification of Junctions in Road Networks ». *GeoInformatica*, Vol. 3, Issue 2, 1999. pp. 185-200.
- Mustière, S., Devogele, T., « Matching networks with different levels of detail ». *GeoInformatica*, Vol.12, n°4, 2008, pp. 435-453.
- Paget, A., Perret, J., Gleyze, J.-F., 2008, La géomatique au service de la caractérisation automatique des réseaux hydrographiques. *Revue en ligne Physio-Géo*, <<http://www.physio-geo.fr/numeros/Volume2/Article%20Pagetperret.pdf>>, volume II, 2008, pp. 147-160
- Plazanet, C., « Modélisation de la géométrie des objets linéaires en vue de leur généralisation cartographique ». *Actes de la 17ième International Cartographic Conference*. Barcelone, 1995, Vol. 1, pp. 486-495
- Richardson D.E., Thomson R.C., « Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks ». *Cartographica*. Volume 33 (1) 1996, pp. 75-83.
- Ruas, A., « Modèle de généralisation de données géographiques à base de contraintes et d'autonomie ». Mémoire de thèse. Laboratoire COGIT (1999).
- Ruas, A., « Le changement de niveau de détail dans la représentation de l'information géographique ». Habilitation à diriger des recherches, Université de Marne-la-Vallée, 2004.
- Perret, J., Ruas, A., Mas, A., « Understanding Urban Dynamics: The Use of Vector Topographic Databases and the Creation of Spatio-Temporal Databases ». *Actes de la 24ème ICC*, Chili, 2009. A paraître.
- Sheeren, D., « L'appariement pour la constitution de bases de données géographiques multi-résolutions. Vers une interprétation des différences de représentation ». *Revue Internationale de Géomatique* Vol 12 n.2/2002 pp 151-168

- Thomson, R.C., Richardson, D., « The "Good Continuation" principle of Perceptual Organization applied to the Generalization of Road Networks », in *proceedings of the 19th ICC*. ICA, Ottawa, Canada (1999).
- Thomson, R., Brooks, R., « Efficient generalisation and abstraction of network data using perceptual grouping ». In *proceedings of the 5th GeoComputation*. University of Greenwich, Kent U.K. (2000)
- Touya, G., « A Road Network Selection Process Based on Data Enrichment and Structure Detection ». In *proceedings of 10th ICA Workshop on Generalisation and Multiple Representation*. Moscow (Russia). 2007a
- Touya, G., « Un processus de sélection du réseau hydrographique, basé sur la détection des structures ». *Le Monde des Cartes*. Revue du Comité Français de Cartographie. Num. 194. Décembre 2007b. pp. 46-56.
- Touya G., Grosso E., Mustière S., Duchêne C., « Generalization or Data Matching for Integration? Insights from Experiments on Databases at Different Levels of Details ». In *Proceedings of the 4th International Conference on Geographic Information Science (GIScience'2006) – Extended Abstracts*. Raubal, Miller, Frank and Goodchild (eds), IFGI Prints, vol.28, pp.373-376.
- Vangenot, C., Parent, C., Spaccapietra, S., « Modelling and manipulating multiple representations of spatial data ». *Proceedings of the 10th International Symposium on Spatial Data Handling*, 2002, pp. 81-93.
- Weibel, R., Dutton, G., « Constraint-Based Automated Map Generalization ». In *proceedings of the 8th International Symposium on Spatial Data Handling*, 1998, pp.214-224.
- Wertheimer, M., « Laws of organization in perceptual forms ». First published as *Untersuchungen zur Lehre von der Gestalt II*, in *Psychologische Forschung*, 4, 301-350. Translation published in Ellis, W. (1938). *A source book of Gestalt psychology* (1923) 71-88