HAL
open science

Une approche basée graphes pour la détection de zones fonctionnelles urbaines

Noudéhouénou Lionel Jaderne Houssou, Jean-Loup Guillaume, Armelle Prigent

To cite this version:

Noudéhouénou Lionel Jaderne Houssou, Jean-Loup Guillaume, Armelle Prigent. Une approche basée graphes pour la détection de zones fonctionnelles urbaines. EGC, Jan 2019, Metz, France. hal-02363996

HAL Id: hal-02363996

https://hal.science/hal-02363996v1

Submitted on 14 Nov 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.

Une approche basée graphes pour la détection de zones fonctionnelles urbaines

Noudéhouénou Lionel Jaderne Houssou*, Jean-Loup Guillaume*
Armelle Prigent*

*L3i, Université de La Rochelle,Avenue Michel Crépeau, 17042 La Rochelle, France
noudehouenou.houssou1,jean-loup.guillaume,armelle.prigent@univ-lr.fr

Résumé. Dans cet article, nous proposons une méthode pour l'identification de zones fonctionnelles, utilisant la détection de communautés dans un graphe de mobilité. Les sommets du graphe correspondent à des unités spatiales, issues du découpage d'une ville suivant le réseau routier. Les arêtes relient des sommets entres lesquels des déplacements sont observés et sont pondérées en fonction du nombre de déplacements et de la distance entre sommets. Notre approche optimise la modularité sur ce réseau pour assurer que les zones fonctionnelles obtenues maximisent les interactions spatiales en leur sein. De plus, nous utilisons les points d'intérêts pour maintenir une hétérogénéité suffisante dans les zones détectées. Nous avons mené des expérimentations avec des trajectoires de taxi et des points d'intérêts de la ville de Porto, afin de montrer la capacité de notre approche à identifier les zones fonctionnelles.

1 Introduction

D'ici 2050, les villes auront 2,5 milliards d'habitants de plus qu'aujourd'hui [1] et plus des deux-tiers de la population mondiale sera urbaine. Cette projection impose de repenser les méthodes de gestion urbaine pour faire face aux challenges qui en découleront. Dans cette optique, une approche consiste à tirer parti des données décrivant les activités des citadins pour découper les villes en zones fonctionnelles, capables de fournir une meilleure compréhension de la structuration spatiale et socio-économique des cités.

En effet, les zones fonctionnelles urbaines sont un formidable outil d'aide à la décision pour la planification urbaine et sont également utiles pour la recommandation de circuits touristiques ou de sites d'implantation de nouveaux commerces. Cependant, délimiter automatiquement des zones fonctionnelles urbaines peut s'avérer complexe car le concept de zone fonctionnelle n'a pas, à l'heure actuelle, de définition harmonisée et les segmentations de villes obtenues dans l'état de l'art sont difficiles à valider comme montré dans Farmer et Fotheringham (2011). Toutefois, cela n'influe pas sur l'intérêt scientifique porté à la délimitation des zones fonctionnelles, au vu de la multiplicité des travaux de recherche traitant du sujet. Lesdits travaux se fondent le plus souvent sur des méthodes de clustering que l'on peut répartir en

La version originale de cet article a été publiée en anglais dans la conférence SAC 2019 (Houssou et al., 2019)

1. https ://www.un.org/development/desa/fr/news/population/2018-world-urbanization-prospects.html

deux grandes familles. D'une part, les approches statistiques, développées par les auteurs qui assimilent les zones fonctionnelles urbaines aux zones résidentielles, commerciales et autres. D'autre part celles à base de graphes qui sont proposées par des chercheurs considérant les zones fonctionnelles urbaines comme des régions marquées par de fortes interactions internes. Nos travaux se positionnent dans le second groupe car l'objectif de notre étude est d'identifier des zones hétérogènes abritant de fortes interactions socio-économiques. En d'autres termes, nous recherchons des villes au sein de la ville.

Dans cet article nous présentons donc une approche à base de graphe pour délimiter les zones fonctionnelles en utilisant à la fois des données de mobilité et des points d'intérêt (POI), sachant que dans les autres approches basées sur les graphes seules les données de mobilité sont exploitées. Le reste de l'article est organisé comme suit : nous donnons plusieurs définitions de la notion de zone fonctionnelle dans la section 2. Nous présentons ensuite plusieurs travaux proches de notre proposition dans la section 3. Puis nous décrivons notre méthodologie pour extraire les zones fonctionnelles dans la section 4 et présentons les résultats obtenus en section 5 avant de conclure dans la section 6.

2 Le concept de zone fonctionnelle

Le problème de la délimitation des zones fonctionnelles est fréquemment et diversement abordé dans la littérature. Cependant, à notre connaissance, aucune défintion consensuelle n'a pu être donnée jusqu'ici à la notion de zone fonctionnelle. Plusieurs propositions coexistent, chacune ayant sa particularité. Le plus souvent elles mettent l'accent sur les interactions socio-économiques au sein des zones fonctionnelles.

Par exemple, l'OCDE (OCDE, 2002) définit une zone fonctionnelle comme une unité territoriale résultant de l'organisation de relations sociales et économiques de sorte que ses frontières ne reflètent pas de particularités géographiques ou historiques. En général, elle est organisée autour d'un ou de plusieurs noeuds de sorte que les zones alentour soient connectées à ce(s) noeud(s) à travers différents systèmes (transports, communications, travail, échanges) (CSIL, 2015). Dans le même ordre d'idées, Karlsson (2007), postule qu'une zone fonctionnelle est caractérisée par une agglomération d'activités et par des infrastructures de transport intra-régionales qui facilitent la mobilité des individus et des produits à l'intérieur de ses frontières.

D'autres auteurs se font plus précis, en restreignant les interactions socio-économiques au cadre du marché du travail et donc aux déplacements travail-maison. Ainsi, selon Antikainen (2005), une zone fonctionnelle est une zone de mobilité domicile-travail. Elle est une agglomération de lieux de travail qui attirent la main d'oeuvre des régions environnantes. Sa qualité la plus importante est sa capacité à dépasser les frontières administratives. Farmer et Fotheringham (2011), abondent dans le même sens en considérant les zones fonctionnelles comme des régions géographiques au sein desquelles les interactions en termes de trajets domicile-travail sont maximisées et entre lesquelles ces interactions sont plutôt minimisées.

Au delà des interactions socio-économiques, une zone fonctionnelle est aussi définie comme un territoire possédant une fonction particulière (zone résidentielle, commerciale, touristique, etc.) qui dépend des activités humaines s'y déroulant. Suivant cette conception, une zone fonctionnelle correspond à une certaine utilisation des terres (Zhi et al., 2014; Gao et al., 2017).

En dépit de leur diversité, ces définitions permettent d'identifier des caractéristiques essentielles concernant les zones fonctionnelles. Premièrement et fondamentalement, elles sont

délimitées de sorte que les interactions socio-économiques sont plus fortes en leur sein qu'entre elles. Deuxièmement, elles sont hétérogènes parce qu'elles regroupent différents types d'activités. Troisièmement, elles ne correspondent pas à un découpage administratif, géographique ou historique. Enfin, elles peuvent être associées à des utilisations spécifiques des terres.

Précisons que lorsque les zones fonctionnelles sont délimitées à l'échelle de centres urbains, elles sont qualifiées de zones fonctionnelles urbaines et c'est précisément l'identification de ce type de zones qui nous intéresse dans cet article. Par ailleurs, du fait des caractéristiques sus-citées, la détermination des zones fonctionnelles impose de disposer de données décrivant des interactions socio-économiques. Dans notre contexte, nous utiliserons des données de mobilité et des POI pour leur identification.

3 Travaux proches

L'identification des zones fonctionnelles urbaines, dans la plupart des travaux récents, rime avec la détermination de l'utilisation des terres. Habituellement, délimiter des zones fonctionnelles urbaines revient à délimiter des zones résidentielles, commerciales, administratives, touristiques, etc. Cette tendance traduit une perception de la zone fonctionnelle urbaine comme étant un espace remplissant une fonction spécifique aux yeux des usagers qui le fréquentent. Plusieurs méthodes, essentiellement statistiques, ont été proposées afin de délimiter automatiquement ce type d'espace. Ainsi, dans Yuan et al. (2012), les auteurs présentent une approche basée sur une régression de mélanges de Dirichlet multinomiaux (DMR pour Dirichlet-Multinominal Regression) pour identifier les zones fonctionnelles de Beijing. Ils utilisent à la fois des données décrivant les déplacements des taxis et les POI dans leur modèle qui est une amélioration du LDA (Latent Dirichlet Allocation). D'autres méthodes utilisant la factorisation de matrices (Wang et al., 2018), l'algorithme Expectation Maximization (Long et Shen, 2015), le clustering hiérarchique (Tu et al., 2018) pour ne citer que celles-là, ont aussi été développées. Elles usent des données provenant des GSM, des trajectoires de taxis, des media sociaux, des carte de bus, d'images satellites, etc.

Toutefois, les zones fonctionnelles urbaines ne sont pas exclusivement associées à l'utilisation des terres dans la littérature. Certains auteurs les conçoivent comme des régions présentant des interactions socio-économiques plus fortes en leur sein que vis à vis de l'extérieur. Nous partageons cette conception car elle correspond à notre sens à la caractéristique fondamentale et traditionnelle de la notion de zone fonctionnelle. Les méthodes développées pour extraire les zones fonctionnelles suivant cette conception présentent une constante. Elles modélisent sous forme de graphe les interactions qui ont lieu dans un territoire puis recherchent des communautés au sein de ce graphe. Ceci s'explique par le fait que premièrement, les graphes permettent de modéliser naturellement les interactions et deuxièmement, la notion de communauté dans un graphe sert à formaliser aisément le concept de zones fonctionnelles. Une communauté étant un ensemble de noeuds plus fortement inter-connectés entre eux qu'avec l'extérieur. Cette formalisation est explicitée dans Farmer et Fotheringham (2011). Les auteurs y proposent une approche pour identifier les zones fonctionnelles, fondée sur l'extraction de communautés dans un graphe obtenu à partir des flux travail-domicile sur tout le territoire Irlandais. Les poids des liens du graphe dépendent à la fois de la distance séparant deux noeuds et du nombre de déplacements observés entre eux. La particularité de leur travail réside dans l'utilisation d'une méthode spectrale pour maximiser la modularité. Pour découvrir les zones fonctionnelles de

Shanghai, Fan et al. (2015) recherchent également des communautés, mais plutôt par le biais de l'algorithme Fast-Newman, dans un graphe de cellules de Voronoï construit à partir des trajectoires de taxis. Demsar et al. (2014) se démarquent des deux travaux précédents en cherchant à détecter des communautés recouvrantes et donc des zones fonctionnelles qui se chevauchent dans la ville de Londres, à partir de données sur la circulation des taxis.

Ces approches à base de graphes n'exploitent que des données d'interaction, plus précisément de mobilité. Elles ne prennent pas en compte la composition socio-économique des territoires étudiés. Or les zones fonctionnelles ne sont pas que des zones de fortes interactions, ce sont aussi des zones hétérogènes rassemblant différents types d'activités. Par conséquent, l'intégration des données socio-économiques peut aider à améliorer la qualité du processus de détermination des zones fonctionnelles. C'est la raison pour laquelle nous proposons une méthode de délimitation des zones fonctionnelles fondée sur l'extraction de communautés et qui enrichit celles déjà existantes en alliant l'utilisation des POI aux données de mobilité. Nous décrivons cette nouvelle approche dans la section suivante.

4 Méthodologie

En nous référant aux définitions de la section 2, l'une des caractéristiques principales des zones fonctionnelles est d'avoir de plus fortes interactions internes que vis à vis de l'extérieur. La détection de telles zones a été formalisée par Farmer et Fotheringham (2011) comme un problème de détection de communautés dans un graphe d'interaction. Problème qui peut être résolu en maximisant la modularité.

La modularité, proposée par Newman (2006), est une mesure évaluant la qualité d'une partition des sommets d'un graphe. Cette mesure utilise le fait qu'une communauté est plus connectée en interne que vers l'extérieur. Elle est formellement définie par l'équation 1 pour un graphe de matrice d'adjacence *A* et une partition *P* des sommets de ce graphe :

$$Q(P) = \frac{1}{2m} \sum_{C \in P} \sum_{i,j \in C} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \tag{1}$$

où *C* est une partie de *P*, *i* et *j* sont deux sommets de *C* avec k_i, k_j leur degré respectif et *m* est le nombre d'arêtes dans le graphe.

L'idée sous-tendant cette formulation consiste à comparer, pour chaque partie, ou communauté, le nombre de liens internes (somme des A_{ij}) avec le nombre de liens attendus dans un modèle de référence (somme des $k_i k_j / 2m$). La qualité est ainsi d'autant meilleure que le nombre d'arêtes observées est supérieur à ce qui était attendu. Pour la modularité le modèle de référence est le modèle configurationnel (Bender et Canfield, 1978) qui préserve uniquement la distribution de degrés du graphe original mais mélange toutes les arêtes. La modularité prend des valeurs comprises entre -0,5 et 1.

Notre méthode s'inspire de celle de Farmer et Fotheringham (2011) et repose également sur la maximisation de la modularité mais (i) nous y intégrons des informations sur les points d'intérêt pour contrôler l'hétérogénéité des zones identifiées et (ii) nous remplaçons la méthode spectrale d'optimisation par une autre plus efficace. Les trois étapes de notre approche sont : segmentation du territoire en unités spatiales, construction des graphes modélisant les interactions entre ces unités et clustering des unités spatiales pour extraire des zones fonctionnelles.

4.1 Segmentation du territoire

Le découpage en grille est une méthode classique de segmentation du territoire très utilisée car simple à implémenter et produisant des zones de surface identique (Wang et al., 2016; Liu et al., 2015; Pei et al., 2014; Liu et al., 2012). Cependant, le découpage qui en résulte ne respecte pas l'aménagement urbain, manque de sémantique (Yuan et al., 2012) et dépend d'un paramètre fixant la taille des cellules.

Afin de pallier ces insuffisances, nous avons préféré diviser l'espace d'étude en zones disjointes en utilisant le réseau routier. Les unités spatiales obtenues sont *a priori* homogènes d'un point de vue socio-économique et évitent les limitations du découpage en grille. Nous utiliserons le terme de zones formelles pour les identifier dans la suite de l'article, comme dans Bednarz et al. (1994).

4.2 Graphe d'interaction

Les interactions entre zones formelles sont modélisées par un graphe de flux, pondéré par le nombre de trajets entre les zones sources et destinations durant une certaine période (en pratique nous ignorons la direction ce qui rend la matrice symétrique). Afin d'obtenir des régions fonctionnelles connexes, les pondérations sont ajustées avec une pondération géographique : les poids sont calculés grâce à une fonction Gaussienne, voir équation 2 telle que proposée par Farmer et Fotheringham (2011) :

$$A_{ij} = W_{ij} \exp(-d_{ij}^2/h^2) \tag{2}$$

où A_{ij} est l'interaction pondérée entre les zones formelles i et j, W_{ij} représente le nombre de trajets entre i et j, d_{ij} est la distance euclidienne entre i et j, h est un paramètre contrôlant la variance de la Gaussienne et donc la compacité des zones.

4.3 Extraction de zones fonctionnelles

L'optimisation de la modularité est un problème NP-difficile (Brandes et al., 2006) mais de nombreuses heuristiques ont été proposées pour trouver rapidement des solutions de bonne qualité. Selon Fortunato (2010), la méthode de Louvain proposée par Blondel et al. (2008) est l'une des méthodes les plus efficaces à la fois en temps et en qualité, c'est la raison pour laquelle nous l'avons sélectionnée pour calculer les partitions.

Afin d'extraire les zones fonctionnelles nous avons donc appliqué la méthode de Louvain sur divers graphes d'interaction obtenus en faisant varier le paramètre h (la variance) de la Gaussienne. Puis, pour chaque partition, nous avons mesuré l'hétérogénéité des régions générées en calculant leur entropie moyenne, comme définie par l'équation 3, grâce à la distribution des points d'intérêt.

$$E_{average} = \frac{1}{|R|} \sum_{R \in P} E_R \tag{3}$$

où E_R est l'entropie (Shannon, 1948) de la région R calculée par :

$$E_R = -\sum_{i \in R} p_i * \log p_i \tag{4}$$

avec p_i la proportion de la i-ème catégorie de points d'intérêts dans la région R.

Notre intuition est que plus une région est hétérogène, plus son entropie sera élevée et plus y devrait y avoir des interactions spatiales en son sein. Enfin, nous sélectionnons la partition qui maximise le produit de la modularité par l'entropie moyenne. Ce produit constitue une règle de sélection multi-critères pour s'assurer que les zones fonctionnelles aient de fortes interactions internes et une hétérogénéité suffisante.

5 Résultats et discussions

5.1 Données et zone d'étude

Notre zone d'étude, Porto, est le centre de la seconde plus grande zone urbaine du Portugal, après Lisbonne. Elle a une population de 237591 habitants (recencement de 2011, http ://mapas.ine.pt) et couvre une surface de 41,42 km2. Cette ancienne ville européenne, également capitale de la région Nord est bordée par le fleuve Douro et l'océan Atlantique.

Porto a un système de transport combinant bus, tramways, métro et taxis. Certains taxis sont équipés de terminaux mobiles capables d'enregistrer leurs positions. Suite à la collecte des données issues de ces taxis, le jeu de données ouvert "Taxi Service Trajectory" a été construit et mis à disposition pour la compétition "ECML/PKDD 15 Kaggle competition". Il comporte les trajets effectués par 442 taxis pendant un an, du 1er juillet 2013 au 30 juin 2014 dans Porto (Moreira-Matias et al., 2013). Chaque trajet comporte un identifiant, l'origine de la demande (envoi par le central, demande à un arrêt ou dans la rue, etc.), un identifiant anonymisé du téléphone du demandeur, l'arrêt de taxi en cas de demande à un arrêt, l'identifiant du taxi, un horodatage, le type de jour (vacances, jour ouvré, week-end), l'absence éventuelle de données GPS, et les coordonnées GPS collectées toutes les 15 secondes.

Le jeu de données contient 1710670 enregistrements. Après un léger nettoyage pour exclure les trajets incomplets, trop courts ou trop longs, il nous reste 1438924 trajectoires. Les trajectoires très courtes peuvent avoir leurs origines et leurs destinations associés à la même zone formelle et les trajectoires très longues peuvent avoir une extrémité en dehors de la zone d'étude. Dans cette étude nous n'utilisons que la source et la destination des trajectoires que nous stockons dans une base de données Postgresql indexée avec l'extension Postgis, utile pour manipuler et effectuer des calculs sur des données spatiales.

Outre ces données, 7710 points d'intérêt ont été collectés depuis openstreetmap [2]. Ces POI sont étiquetés par des types de base (distributeur de billets, banque, église, hotel, etc.) que nous avons regroupés en dix catégories plus générales pour en réduire le nombre. Ces catégories sont : Logement, Lieux de travail et services publics, Alimentation et loisirs, Tourisme, Transports, Business et industries, Education, Santé, Espaces verts, Lieux de culte.

Enfin, les zones formelles sont extraites de l'atlas urbain de 2012 disponible en ligne [3]. Il s'agit de données vectorielles avec des informations cadastrales. Le processus de segmentation fournit 2453 zones formelles qui sont représentées sur la figure 1.

2. https ://www.openstreetmap.org
3. http ://mapas.dgterritorio.pt/atom-dgt/CDG_atlasurbano2012_Continente_Atom.xml

FIG. 1 – *Zones formelles de Porto.*

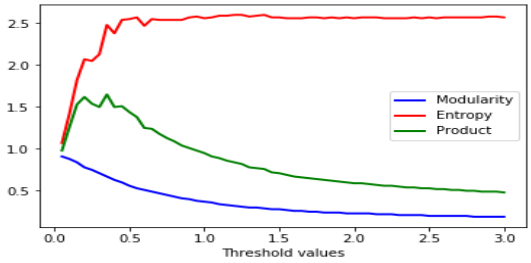

FIG. 2 – *Modularité et entropie moyenne en fonction de la variance h de la Gaussienne.*

5.2 Experimentations et résultats

Nos expériences ont été menées sur un ordinateur portable HP Zbook, possédant un processeur octocœur Intel i7-6700HQ ayant une fréquence de 8 * 2.60Ghz et 16Gb de Ram. Les opérations de SIG sont effectuées avec QGIS sous Ubuntu 16.04 LTS.

A partir des zones formelles et des données origines-destinations, nous avons construit des graphes d'interaction pour des valeurs de la variance de la Gaussienne (le paramètre h de l'équation 2) comprises entre 0,05 et 3 avec un pas de 0,05. Les régions n'ayant aucune interaction sont supprimées lors de la phase de construction du graphe. La méthode de Louvain est ensuite exécutée 10 fois sur chaque graphe et la partition de plus forte modularité est conservée. L'entropie moyenne est ensuite calculée ainsi que le produit modularité par entropie moyenne. Les résultats sont présentés figure 2. Notons que la condition d'arrêt de notre expérimentation est la convergence de l'entropie moyenne. Or d'après la figure 2 l'entropie moyenne stagne pour des valeurs de variance comprises entre 1,5 et 3. Ce qui explique que nous n'ayons pas effectué d'expérimentations au delà.

La variance qui maximise le produit entre entropie moyenne et modularité est $h = 0,35$. La partition correspondante contient 11 communautés qui correspondent aux zones fonctionnelles illustrées figure 3. Les figures 4 et 5 montrent deux exemples avec des variances valant respectivement $0,1$ et 2. On peut noter une corrélation claire entre la taille des communautés et la variance même si nous n'avons pas formalisé cette corrélation.

FIG. 3 – *Zones fonctionnelles obtenues avec une variance* $h = 0,35$.

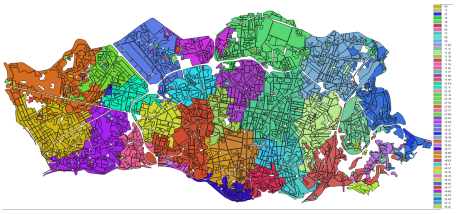

FIG. 4 – *Zones fonctionnelles obtenues avec une variance* $h = 0,1$.

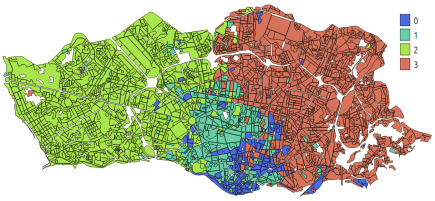

FIG. 5 – *Zones fonctionnelles obtenues avec une variance* $h = 2$.

Après comparaison, les zones fonctionnelles obtenues ne correspondent pas aux frontières administratives des districts de Porto représentées figure 6 même si on peut noter quelques similarités. Cela confirme que la ville a une structure cachée qui n'est accessible que via l'exploitation des données de mobilité. La diversité des activités de chaque zone est présentée figure 7 (les couleurs identifiant les zones et les POI ne sont pas corrélées). La figure 7 montre que les communautés ont une bonne variété de types de POI et que les zones résidentielles sont dominantes dans la majorité des communautés. Cette situation est normale étant donné que fournir des logements est l'une des principales, sinon la principale, fonction d'une ville.

Bien que les communautés détectées soient de grands blocs contigus, il arrive que certaines de leurs zones formelles se retrouvent en dehors de leurs frontières. Cela vient du fait que quelques zones formelles géographiquement éloignées restent fortement interconnectées même en pénalisant grandement les distances élevées avec le calcul de l'équation 2. Ce phénomène peut être géré en assignant aux zones formelles l'identifiant de la communauté dans laquelle elles sont incluses ou, au contraire, il est possible de les laisser en l'état pour ne pas perdre l'information de la forte connectivité.

D'après les expérimentations de Farmer et Fotheringham (2011), la variance h devrait être calculée de manière adaptative pour chaque origine à partir des distances des trajets partant de cette origine et aboutissant à d'autres zones formelles. Néanmoins, ce calcul génère des zones fonctionnelles très dispersées comme illustré figure 8. La méthode de Farmer et Fotheringham (2011) s'avère donc assez inefficace pour la détection de zones fonctionnelles à l'échelle d'une ville. Comme solution, nous avons choisi une approche globale plutôt que locale. Autrement

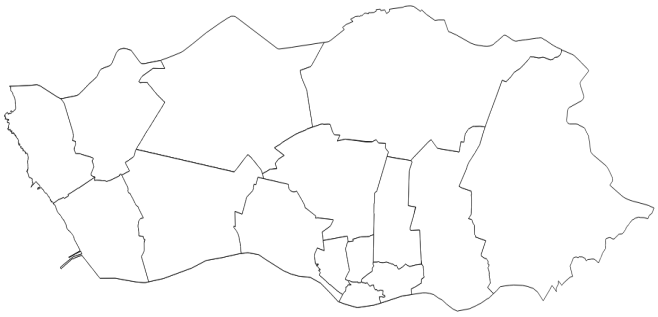

FIG. 6 – *Districts de Porto.*

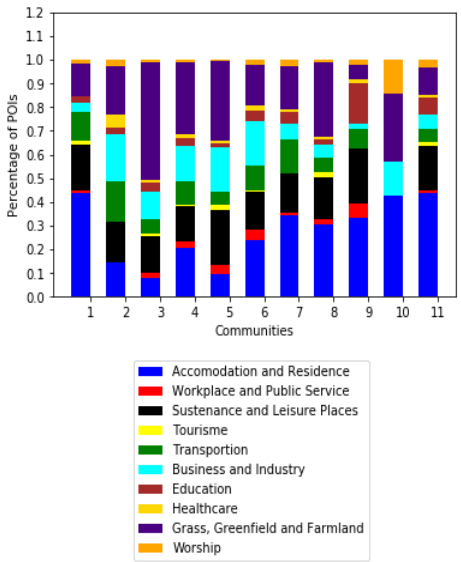

FIG. 7 – *Distribution de points d'intérêts par communauté.*

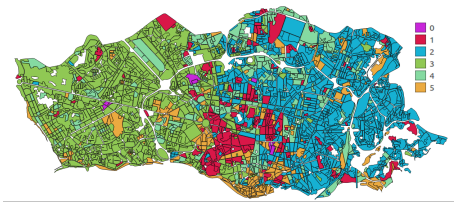

FIG. 8 – *Zones fonctionnelles avec une évaluation automatique de la variance pour chaque région.*

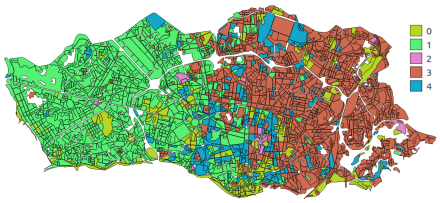

FIG. 9 – *Zones fonctionnelles avec une évaluation automatique de la variance pour toutes les régions.*

dit, nous déterminons la variance h pour l'ensemble des zones formelles en ajustant la distribution des distances avec une loi gaussienne. Cependant, ce mode de calcul génère également des zones fonctionnelles très dispersées comme illustré figure 9. En raison des résultats de mauvaise qualité, obtenus avec ces deux modes de sélection automatique de la variance, nous ne sommes pas en mesure de comparer notre modèle à celui proposé par Farmer et Fotheringham (2011) et qui nous sert de modèle de référence.

6 Conclusions et perspectives

Cet article propose une méthode à base de graphes pour délimiter les zones fonctionnelles d'une ville. Cette méthode consiste à diviser la ville en zones formelles à partir du réseau routier puis à construire un graphe d'interaction entre ces zones en utilisant les trajectoires de taxis. Les noeuds du graphe correspondent aux zones formelles et deux noeuds sont connectés si il existe un ou des trajets de taxis allant d'une zone à l'autre. Les arêtes sont pondérées en fonction du nombre de trajets et de la distance entre zones formelles. L'utilisation d'une fonction à décroissance gaussienne permet de pénaliser les longues distances et d'obtenir un ensemble de graphes en modifiant la valeur de la variance. Les graphes obtenus sont ensuite décomposés en communautés maximisant la modularité. Chaque communauté correspond à une zone fonctionnelle. Le meilleur découpage en zones fonctionnelles est identifié en utilisant l'information apportée par l'entropie des points d'intérêts au sein des communautés. Toutes les expériences ont été menées sur la ville de Porto grâce à des trajectoires de taxis enregistrées pendant un an et des POI.

Ce travail peut être approfondi en étudiant la stabilité des zones détectées au cours du temps. En particulier, les données pourraient être divisées en sous-ensembles couvrant différentes périodes et les zones fonctionnelles identifiés pour chaque sous-ensemble. On déterminerait ensuite des coeurs de communautés (Gfeller et al., 2005) pour l'ensemble des partitions obtenues. Par ailleurs, il peut également être pertinent de calculer des communautés recouvrantes pour obtenir un découpage plus réaliste qui ne soit pas une partition stricte. En outre, l'utilisation des trajectoires est pour l'instant minimale étant donné que seules les zones de départ et d'arrivée sont exploitées. Les zones traversées pourraient aider à affiner la détection des zones fonctionnelles. Enfin, la dimension temporelle peut également apporter une information forte pour la pondération des arêtes du graphe.

Références

Antikainen, J. (2005). The concept of functional urban area. findings of the espon project. *Informationen zur Raumentwicklung 7*, 447–452.

Bednarz, S. et al. (1994). *Geography for Life : National Geography Standards*. Distributed by ERIC Clearinghouse (Washington, D.C.).

Bender, E. et E. Canfield (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A 24*(3), 296 – 307.

Blondel, V., J.-L. Guillaume, R. Lambiotte, et E. Lefebvre (2008). Fast unfolding of communities in large networks. *J. Stat. Mech*, P10008.

Brandes, U., D. Delling, M. Gaertler, R. Goerke, M. Hoefer, Z. Nikoloski, et D. Wagner (2006). Maximizing modularity is hard. cite arxiv :physics/0608255 Comment : 10 pages, 1 figure.

CSIL (2015). *Territorial Agenda 2020 put in practice. Enhancing the efficiency and effectiveness of Cohesion Policy by a place-based approach*. Publications Office of the European Union.

Demsar, U., J. Reades, E. Manley, et J. Batty (2014). Edge-based communities for identification of functional regions in a taxi flow network. In *Extended Abstract Proceedings of the GIScience 2014*, pp. 55–60.

Fan, K., D. Zhang, Y. Wang, et S. Zhao (2015). Discovering urban social functional regions using taxi trajectories. In *Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing*, pp. 356–359.

Farmer, C. et A. Fotheringham (2011). Network-based functional regions. *Environment and Planning A : Economy and Space 43*(11), 2723–2741.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports 486*(3-5), 75 – 174.

Gao, S., K. Janowicz, et H. Couclelis (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS 21*(3), 446–467.

Gfeller, D., J.-C. Chappelier, et P. D. L. Rios (2005). Finding instabilities in the community structure of complex networks. *Phys. Rev. E 72*, 056135.

Houssou, N., J.-L. Guillaume, et A. Prigent (2019). A graph based approach for functional urban areas delineation. In *Proceedings of the 34rd Annual ACM Symposium on Applied Computing, SAC 2019, Limassol, Cyprus, April 08-12, 2019*.

Karlsson, C. (2007). Clusters, Functional Regions and Cluster Policies. Working Paper Series in Economics and Institutions of Innovation 84, Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies.

Liu, X., L. Gong, Y. Gong, et Y. Liu (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography 43*, 78 – 90.

Liu, Y., F. Wang, Y. Xiao, et S. Gao (2012). Urban land uses and traffic 'source-sink areas' : Evidence from gps-enabled taxi data in shanghai. *Landscape and Urban Planning 106*(1), 73 – 87.

Long, Y. et Z. Shen (2015). *Discovering Functional Zones Using Bus Smart Card Data and Points of Interest in Beijing*, pp. 193–217. Cham : Springer International Publishing.

Moreira-Matias, L., J. Gama, M. Ferreira, J. Mendes-Moreira, et L. Damas (2013). Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems 14*(3), 1393–1402.

Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences 103*(23), 8577–8582.

OCDE (2002). *Redefining Territories : The Functional Regions*. Éditions OCDE, Paris.

Pei, T., S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, et C. Zhou (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science 28*(9), 1988–2007.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423.

Tu, W., Z. Hu, L. Li, J. Cao, J. Jiang, Q. Li, et Q. Li (2018). Portraying urban functional zones by coupling remote sensing imagery and human sensing data. *Remote Sensing 10*(1).

Wang, Y., Y. Gu, M. Dou, et M. Qiao (2018). Using spatial semantics and interactions to identify urban functional regions. *ISPRS International Journal of Geo-Information 7*(4).

Wang, Y., T. Wang, M.-H. Tsou, H. Li, W. Jiang, et F. Guo (2016). Mapping dynamic urban land use patterns with crowdsourced geo-tagged social media (sina-weibo) and commercial points of interest collections in beijing, china. *Sustainability 8*(11).

Yuan, J., Y. Zheng, et X. Xie (2012). Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, New York, NY, USA, pp. 186–194. ACM.

Zhi, Y., Y. Liu, S. Wang, M. Deng, J. Gao, et H. Li (2014). Urban spatial-temporal activity structures : a new approach to inferring the intra-urban functional regions via social media check-in data. *CoRR abs/1412.7253*.

Summary

In this paper, we presented a method for the structural analysis of a city, specifically for the determination of its functional areas, based on communities detection in graph. The nodes of the graph correspond to geographical units resulting from a cartographic division of the city according to the road network. The edges are weighted using a Gaussian distance-decay function and the amount of spatial interactions between nodes. Our approach optimize the modularity to ensure that the functional areas detected have strong interactions within their borders but lower interactions outside. Moreover, it leverages on POIs' entropy to maintain a good socioeconomic heterogeneity in the detected areas. We conducted experiments using taxi trips and POIs datasets from the city of Porto, as a study case, to demonstrate the ability of our method to portray functional areas while including spatial and socioeconomic dynamics.