

# Identifier, détecter et localiser les centralités économiques : une proposition inspirée de l'algorithme DBSCAN

## Identifying, detecting and locating economic centralities: A proposal inspired by the dbscan algorithm

Jean Dubé, Philippe Bilodeau, M. ATDR, Gabriel Sylvain-Nolet, M. ATDR,  
Claudiel Keuneng, M. ATDR, Oussema Amir Boumankar, M. ATDR et Joé Dufour,  
M. ATDR

Volume 45, numéro 1, 2022

Soumis : 2021-10-05

Accepté : 2021-12-07

URI : <https://id.erudit.org/iderudit/1088302ar>

DOI : <https://doi.org/10.7202/1088302ar>

[Aller au sommaire du numéro](#)

### Éditeur(s)

Canadian Regional Science Association / Association canadienne des sciences régionales

### ISSN

0705-4580 (imprimé)

1925-2218 (numérique)

[Découvrir la revue](#)

### Citer cet article

Dubé, J., Bilodeau, P., Sylvain-Nolet, G., Keuneng, C., Boumankar, O. A. & Dufour, J. (2022). Identifier, détecter et localiser les centralités économiques : une proposition inspirée de l'algorithme DBSCAN. *Canadian Journal of Regional Science / Revue canadienne des sciences régionales*, 45(1), 48–58. <https://doi.org/10.7202/1088302ar>

### Résumé de l'article

En sciences régionales, la notion de centre(s) revêt une importance capitale pour plusieurs modèles théoriques et cadres conceptuels. Les dynamiques urbaines et locales ont conduit à une multiplication des centres rendant le polycentrisme la norme dans la plupart des aires métropolitaines. Or, la localisation du centre (ou des centres) est souvent postulée comme exogène, déterminée à l'avance. Elle s'avère souvent un intrant nécessaire afin d'identifier les multiples centres. Cette note de recherche propose de développer un algorithme permettant d'identifier, de détecter et de localiser les différents centres à partir d'une typologie issue de critères économiques (unités de logements et commerciales). Afin de présenter son implémentation, deux applications fictives sont développées. Une première sur une ville monocentrique, et une seconde sur une ville polycentrique. Une application empirique permet d'identifier et de localiser les centralités de 31 régions métropolitaines de recensement (RMR) et agglomérations de recensement (AR) du Québec à partir d'information sur les unités d'évaluation contenues dans les rôles d'évaluations municipaux.

# IDENTIFIER, DÉTECTER ET LOCALISER LES CENTRALITÉS ÉCONOMIQUES: UNE PROPOSITION INSPIRÉE DE L'ALGORITHME DBSCAN

Jean Dubé, Philippe Bilodeau, Gabriel Sylvain-Nolet, Claudel Keuneng,  
Oussema Amir Boumankar & Joé Dufour

**Jean Dubé\***  
Université Laval  
2325 rue des Bibliothèques  
Pavillon Félix-Antoine-Savard  
Québec, Québec, Canada, G1V 0A6  
[jean.dube@esad.ulaval.ca](mailto:jean.dube@esad.ulaval.ca)

**Philippe Bilodeau** M. ATDR  
Université Laval

**Gabriel Sylvain-Nolet** M. ATDR  
Université Laval

**Claudel Keuneng** M. ATDR  
Université Laval

**Oussema Amir Boumankar** M. ATDR  
Université Laval

**Joé Dufour** M. ATDR  
Université Laval

Soumis : 2021-10-05

Accepté : 2021-12-07

**Résumé :** En sciences régionales, la notion de centre(s) revêt une importance capitale pour plusieurs modèles théoriques et cadres conceptuels. Les dynamiques urbaines et locales ont conduit à une multiplication des centres rendant le polycentrisme la norme dans la plupart des aires métropolitaines. Or, la localisation du centre (ou des centres) est souvent postulée comme exogène, déterminée à l'avance. Elle s'avère souvent un intrant nécessaire afin d'identifier les multiples centres. Cette note de recherche propose de développer un algorithme permettant d'identifier, de détecter et de localiser les différents centres à partir d'une typologie issue de critères économiques (unités de logements et commerciales). Afin de présenter son implémentation, deux applications fictives sont développées. Une première sur une ville monocentrique, et une seconde sur une ville polycentrique. Une application empirique permet d'identifier et de localiser les centralités de 31 régions métropolitaines de recensement (RMR) et agglomérations de recensement (AR) du Québec à partir d'information sur les unités d'évaluation contenues dans les rôles d'évaluations municipaux.

**Mots clés :** Centralités économiques; Algorithme DBSCAN; Analyse spatiale

**Identifying, detecting and locating economic centralities: a proposal inspired by the dbscan algorithm**

**Abstract:** In regional sciences, the notion of center(s) is of key importance for several theoretical models and conceptual frameworks. Urban and local dynamics have led to a multiplication of centers, making polycentrism the norm in most metropolitan areas. Yet the location of the center (or centers) is often postulated as exogenous, predetermined. It is often a necessary input for identifying multiple centers. This research note proposes to develop an algorithm for identifying, detecting and locating different centers based on a typology derived from economic criteria (housing and commercial units). In order to present its implementation, two fictitious applications are developed. The first is based on a monocentric city, and the second on a polycentric city. An empirical application identifies and locates the centralities of 31 census metropolitan areas (CMAs) and census agglomerations (CAs) in Quebec, based on information on assessment units contained in municipal assessment rolls.

**Keywords:** Economic centralities; DBSCAN algorithm; spatial analysis

\* Cette recherche est en partie financée par le Conseil de recherche en sciences humaines (CRSH) du Canada.

## INTRODUCTION

En sciences régionales, les théories de localisation font largement appel à une variable clé : la localisation du centre, ou du marché. C'est notamment le cas du célèbre modèle de von Thünen (1826), mais aussi des modèles de Weber (1929), de Hotelling (1929), de Hoover (1948) ou encore de celui développé par Alonso (1964), Muth (1967) et Mills (1969) (souvent désigné par le modèle AMM). À cela s'ajoute les modèles de localisation en sociologie urbaine (Grafmeyer & Joseph, 2004).

Concept clé dans les théories en sciences régionales, la centralité est traitée de manière différente selon l'angle dont on l'aborde. Comme le notait déjà Castells (1972), les centralités sont diverses : centralité économique (Polèse et al., 2015; Duranton, 1999), politique (Monnet, 2000), sociale (Ascher, 1999), urbanistique (Gehl, 2012), historique (Halbert, 2006), ou encore selon son accessibilité (Claval, 1981).

Le centre, qui a depuis évolué dans son rôle et ses fonctions, est souvent désigné par le terme *central business district* (CBD – centre des affaires). Le CBD est souvent abordé de manière exogène : sa localisation étant fixée et connue à l'avance. Sans reprendre les nombreuses définitions de la centralité ou les débats autour de cette notion (voir notamment Gaschet et Lacour, 2002 ou Bourdeau-Lepage et al., 2009), cette note de recherche propose un algorithme permettant d'identifier, de détecter et de localiser les centralités à partir de micro-données spatiales.

Cet algorithme permet d'intégrer les derniers développements en lien avec l'utilisation des méga-données et la disponibilité des données administratives. Les résultats issus de l'algorithme, essentiellement descriptifs, revêtent une utilité certaine pour les chercheurs en sciences régionales. Ils peuvent, notamment, être mobilisés comme intrant dans des analyses statistiques. Le principal avantage de l'algorithme est de proposer une méthode standardisée afin d'identifier, de détecter et de localiser les centralités à partir de micro-données administratives sans pour autant limiter l'analyse à l'utilisation de polygones spécifiques, ce qui en constitue une originalité.

Afin de démontrer la faisabilité de cet algorithme, un code sur le logiciel Stata est présenté et expliqué étape par étape. L'algorithme est d'abord présenté de manière générale et appliqué sur deux cas fictifs : un premier de type monocentrique, et un second de type polycentrique. Une application reposant sur les régions métropolitaines de recensement (RMR) et les aires de recensement (AR) du Québec est ensuite développée. L'algorithme peut facilement être bonifié afin d'y inclure des dimensions (quantitatives) supplémentaires afin de considérer des centralités autres qu'économiques.

La note est divisée en six sections. La première section propose une brève revue des méthodes d'identification des centralités économiques. La seconde section propose de présenter une nouvelle approche d'identification, de détection et de localisation des centres à partir d'une approche d'un algorithme de type DBSCAN (Ester et al., 1996). La troisième section présente de manière formelle la méthodologie d'implémentation de l'algorithme. La quatrième section propose deux applications sur des données fictives afin de présenter formellement le code (l'algorithme) et assurer une répliquabilité. La cinquième section propose d'appliquer l'algorithme sur les régions métropolitaines de recensement (RMR) et les agglomérations de recensement (AR) du Québec en utilisant les informations des rôles d'évaluation pour l'année 2020. Une courte conclusion clôture la note.

<sup>1</sup> Plusieurs approches théoriques suggèrent même une forme précise pour cette relation : celle de l'exponentielle négative de la distance (ou  $\exp(-dc)$ , où  $dc$  représente la distance au centre). À noter que cette idée est également reprise dans la construction des matrices de pondérations spatiales (et spatio-temporelles) en économétrie spatiale (Dubé et Legros, 2014).

## LES MÉTHODES EXISTANTES

Dans la littérature, la centralité, ou plutôt les centralités, mobilise un concept de « proximité » à diverses composantes telles que : i) le marché du travail (localisation de l'emploi); ii) la distribution spatiale des populations; iii) l'aménagement urbain; iv) la présence d'infrastructures particulières; v) le marché immobilier; ou encore vi) la présence d'externalités, d'économies d'agglomération ou d'aménités (Tableau 1).

**Tableau 1.** Critères de centralités

Type de centralité	Auteurs	Critères
<b>Économique</b>	Murphy et Vance (1954)	Valeur du sol
	McDonald (1987)	Spécialisation des activités
	Guiliano & Small (1991)	Hauteur des bâtiments
	Polèse et al. (2015)	Commerces rares
	Duranton, (1997)	Absence d'activités non-centrales
	Fujita et Thisse (1997)	Présence d'interactions (marchandes ou non)
	Storper & Venables (2004)	Mixité des activités
<b>Urbanistique</b>	Beaujeu-Garnier (1967)	Hauteur des bâtiments
	Gauthiez (2003)	Forte continuité de tissu urbain
	Ghel (2012)	Forte densité du bâti
	Allain (2004)	Vide structurant
	Chalas (2010)	Mixité des activités
<b>Sociologique</b>	Lebrun (2004)	Mixité sociale
	Roncayolo (1990)	Dimension symbolique
	Devisme (2005)	Urbanité contemporaine
	Lévy (1994)	
	Monnet (2000)	
<b>Politique</b>	Bourdeau-Lepage et al. (2009)	Localisation des pouvoirs publics
	Gaschet et Lacour (2002)	Localisation des institutions
	Monnet (2000)	Commandement
	Ascher (1999)	
	Sassen (1991)	

Dans son livre *Fundamentals of Urban Economics*, McDonald (1997) définit dès le premier paragraphe ce que les économistes considèrent comme une aire urbaine, corrélat important de la centralité. Deux caractéristiques fondamentales y sont soulignées : 1) une forte densité de population par rapport à son environnement immédiat; et 2) une population plus grande qu'un certain seuil minimum. Le point central correspond à l'endroit où la densité de population atteint un sommet (*peak*). La densité déclinerait ensuite avec l'éloignement par rapport à ce point.<sup>1</sup>

La littérature empirique dominante propose d'identifier les centralités économiques à partir de quatre approches dominantes. Ces approches ont une base empirique commune : le recours aux données



d'emploi ou de population (McDonald, 1987). Certaines applications proposent l'utilisation du ratio emploi/résident comme indicateur de centralité.

La première approche consiste à utiliser des valeurs seuils (densité – emploi ou population) minimales et/ou un nombre minimum, au-delà desquelles les secteurs qui satisfont les critères sont identifiés comme des centralités (Giuliano & Small, 1991; McMillen & McDonald, 1990, 1998; Bogart & Ferry, 1999; Cervero & Wu, 1997). Certaines variantes proposent également d'intégrer les sites contigus aux centralités dans la définition des centres.

La seconde approche repose sur l'utilisation de modèles économétriques. L'unité d'analyse est un polygone correspondant à une unité de recensement, la variable dépendante est l'emploi ou la population par secteur, alors que la variable indépendante est la distance au centre principal. La détection repose sur le postulat d'une structure polycentrique qui prend appui sur un centre historique, dans sa vision monocentrique, jugé « exogène ». Les modèles reposent sur des techniques de lissage spatial, qui permettent d'identifier des « pics » de densités (Craig & Ng, 2001), ou encore sur l'identification de valeurs aberrantes (concentrations des résidus) (Small & Song, 1993; McDonald & Prather, 1994; McMillen, 2001; McMillen & Smith, 2003; McMillen & McDonald, 1997; Redfern, 2007).

La troisième approche mobilise l'analyse exploratoire des données spatiales et l'utilisation de statistiques spatiales. Ces statistiques descriptives permettent d'identifier des regroupements de valeurs dans l'espace et d'en tester la significativité. Il est possible de recourir aux indices locaux d'associations spatiales (ou LISA – Anselin, 1995) ou encore aux méthodes de détections de regroupement (*cluster analysis*; Cressie, 1991; Kuldorf, 1997). Ces méthodes permettent de cartographier les concentrations spatiales sur la base d'une variable préalablement identifiée, comme l'emploi ou la population (Griffith & Wong, 2007; Baumont et al., 2004; Guillain et al., 2006; Guillain & Le Gallo, 2010). Une approche complémentaire consiste à tirer profit de matrices de pondérations spatiales pour identifier les polygones ou cellules voisines des unités contenant un certain seuil d'emplois (McMillen, 2003).

Finalement, la dernière approche repose sur l'analyse qualitative et sur les opinions et expertises locales. Il s'agit de tirer profit des connaissances fines des intervenants locaux afin d'identifier les centralités et leurs limites spatiales (Bingham & Kimble, 1995). Cette approche a l'avantage de tirer profit des connaissances tacites, mais a le désavantage d'être difficilement généralisable ou même exploitable à grande échelle.

On peut alors qualifier de centralités primaires les polygones qui respectent les valeurs cibles, et de centralités secondaires les polygones contigus.

## UN ALGORITHME POUR IDENTIFIER, DÉTECTER ET LOCALISER LES CENTRES

La plupart des approches quantitatives reposent sur une définition a priori des limites d'unités spatiales (ou polygones). Le recours à ces approches fait naturellement émerger les problèmes de céoupage et d'agrégation spatiale des entités spatiales (ou *Modifiable Areal Unit Problem* (MAUP) – Openshaw, 1984). La récente disponibilité des micro-données et des méga-données permet d'envisager une approche qui ne nécessite pas de fixer des limites spatiales a priori. Elles permettent une sensibilité plus fine des réalités locales, mais aussi de tester formellement le MAUP en faisant changer le choix du rayon et en comparant les résultats obtenus (Dubé & Brunelle, 2014; Gbenyo & Dubé, 2019).

Le recours aux micro-données spatiales (Dubé & Legros, 2014) et à l'algorithme de type *Density-Based Spatial Clustering of Applications with Noise* (ou DBSCAN – Galan, 2019; Khan et al., 2014) permet de tirer profit de la distribution spatiale d'un ensemble de points (*point pattern analysis*). En ce sens, l'algorithme relève de la troisième approche, mais prend néanmoins appui sur la première approche en utilisant des critères spécifiques sur des variables identifiées à l'avance. Elle s'inspire également de la seconde approche puisqu'elle tient compte de la répartition spatiale exacte des unités d'évaluation, ce qui permet d'obtenir une forme de lissage spatial.

Plus spécifiquement, l'algorithme consiste essentiellement à vérifier si, pour un point donné, les critères fixés sur le nombre d'unités de logement et d'unités de commerce sont satisfaits. Si un point respecte les critères fixés, il est alors considéré comme faisant partie d'une centralité. Si ce n'est pas cas, le point peut être classé comme appartenant à une centralité s'il se trouve « proche » d'un point faisant partie d'une centralité. Autrement, le point est simplement considéré comme hors d'une centralité.

En sciences régionales, la notion de regroupement (ou *clusters*) est particulièrement mobilisée afin de construire des indicateurs de concentrations locales (Dubé & Brunelle, 2014; Kopczewska, 2018) ou encore de développer des statistiques et tests de détection des concentrations spatiales, incluant le recours à l'algorithme DBSCAN (Kopczewska et al., 2017). L'algorithme de type DBSCAN, ou du moins une forme similaire, a récemment été mobilisé par les travaux de De Bellon et al. (2019) et de Rozenfeld et al. (2011), qui utilisent des informations sur des pixels afin d'identifier les concentrations spatiales. Il a aussi été mobilisé par les travaux de Arribas-Bel et al. (2019) qui prennent appui sur la localisation des bâtiments et leurs caractéristiques. Dans le dernier cas, les auteurs proposent également un test de permutation pour juger de la significativité du regroupement spatial.

L'algorithme présenté est inspiré de ces récentes applications et permet de décomposer les centralités à partir d'une typologie préalablement identifiée.

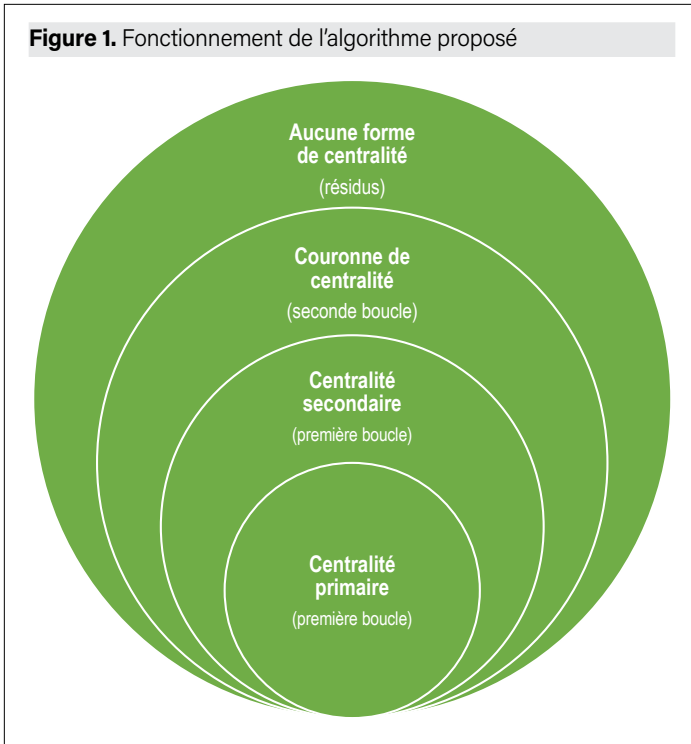
## APPROCHE MÉTHODOLOGIQUE

L'algorithme permet d'identifier, de détecter et de localiser les types de centres (économiques) à partir de trois informations : i) la localisation exacte d'un point  $i$  (les coordonnées  $x$  et  $y$  – noté  $x_{C_i}$  et  $y_{C_i}$ ); ii) le nombre d'unités de logements déclarés à cette localisation ( $u_{L_i}$ ); et iii) le nombre d'unités commerciales ( $u_{C_i}$ ) déclarées à cette localisation. Chaque localisation (ou point) peut contenir aucune, une ou plusieurs unités (logements et commerce).

L'algorithme permet, sur la base de critères fixés sur les unités (logements et commerces), de produire une typologie en quatre catégories (Figure 1). Afin de distinguer les centralités selon leur type, deux critères sont développés : une centralité primaire ( $\tau_1$ ) et une centralité secondaire ( $\tau_2$ ). De la même manière, ces critères sont utilisés afin de déduire les critères de centralités sur les commerces ( $\pi_1$  et  $\pi_2$ ).

Pour un point  $i$  donné, il s'agit, dans une distance critique  $d_C$  fixée, de compter : i) le nombre total d'unités de logement ( $\bar{u}_{L_i} = \sum_j u_{L_j}$  si  $d_{ij} \leq d_C$ , où  $d_{ij}$  représente la distance entre le point  $i$  et le point  $j$ ); et ii) le nombre total d'unités de commerce ( $\bar{u}_{C_i} = \sum_j u_{C_j}$  si  $d_{ij} \leq d_C$ ). Sur la base de ces décomptes, le point est ensuite classé en fonction de : i) la densité résidentielle ( $\tau$ ); et ii) la densité commerciale ( $\pi$ ). Afin de simplifier, la densité commerciale est exprimée en fonction de la densité résidentielle,  $\pi = \theta\tau$ , où  $\theta$  est le rapport entre le nombre de commerces et de logements.

**Figure 1.** Fonctionnement de l'algorithme proposé



Un point  $i$  est qualifié de :

- **centralité primaire** si  $\bar{u}_{Li} \geq \tau_1$  et  $\bar{u}_{Ci} \geq \pi_1$  (ou de manière équivalente, si  $\bar{u}_{Ci} \geq \theta\tau_1$ ).
- **centralité secondaire** si  $\bar{u}_{Li} \geq \tau_2$  et  $\bar{u}_{Ci} \geq \pi_2$  (ou si  $\bar{u}_{Ci} \geq \theta\tau_2$ ) et qu'il n'est pas qualifié de centralité primaire.
- **couronne de centralité** si les critères précédents ne sont pas respectés, mais qu'il est néanmoins situé à une distance inférieure à une certaine valeur critique,  $d_C$ , ( $d_{ij} \leq d_C$ ) d'un point  $j$  qualifié de centralité (primaire ou secondaire).
- **aucune forme de centralité** si le point  $i$  ne se qualifie d'aucun des critères précédents.

La catégorie résiduelle marque donc la non-appartenance à une forme de centralité.<sup>2</sup>

L'application est relativement directe, mais repose sur certains paramètres cruciaux, soit  $\tau_1$ ,  $\tau_2$  et  $\theta$ . Bien que ces valeurs soient modulables selon le type de milieu à l'étude, il est possible de proposer certaines valeurs cibles. Simard (2012) souligne l'importance du seuil de 1 000 habitants par kilomètres carrés afin de définir une aire urbaine au Canada (voir aussi OCDE, 2012). Il est donc réaliste de fixer la valeur de  $\tau_1 = 1000$  pour un rayon ( $d_C$ ) de 565 mètres. De la même manière, Barnes (2001) suggère qu'un milieu urbanisé renferme un emploi pour 5 résidents. On peut donc fixer  $\theta = 1/5$ . Finalement, et

de manière un peu plus subjective, il est possible de fixer la valeur de  $\tau_2 = \tau_1/4 = 250$ . Les centralités secondaires sont donc fixées au quart de la valeur cible pour les centralités primaires. Ces valeurs seuils sont en ligne directe avec les grandes lignes proposées par la classification des centres de populations et des régions rurales de Statistique Canada (2016).

## DEUX APPLICATIONS FICTIVES

Afin de démontrer le fonctionnement de l'algorithme de détection et d'identification des centralités économiques, cette section propose deux exemples fictifs : un premier basé sur une structure monocentrique, et un second basé sur une structure polycentrique. L'algorithme est développé avec le logiciel Stata, mais peut facilement être généralisé sous n'importe quelle plateforme ou logiciel.<sup>3</sup>

Avant tout chose, une étape préliminaire est nécessaire afin de préparer le logiciel (ligne 1) et d'assurer une répliquabilité des coordonnées (ligne 2).

1. `clear all`
2. `set seed 180179`

### Le cas monocentrique

Pour fin de l'exercice, considérons une ville fictive composée de cent unités d'évaluation (ligne 3).

3. `set obs 100`

La première étape consiste à générer la localisation précise des unités individuelles. Ces localisations sont simulées à partir de couple de points représentant les coordonnées géographiques  $x$  et  $y$ . Les coordonnées géographiques  $x$  (ligne 4) et  $y$  (ligne 5) sont simulées sur la base de distributions de probabilité. Il est également intéressant de créer un identifiant unique pour chacun des points (ou paires de coordonnées) (ligne 6).

4. `quietly generate x = rexponential(3)`
5. `quietly generate y = rnormal(3)`
6. `quietly generate id = _n`

La seconde étape consiste à définir les caractéristiques des points. Il s'agit d'identifier les bâtiments voués aux activités commerciales (ligne 7), ainsi que de déterminer le nombre d'unités (logements et commerces) dans chacun des bâtiments (ligne 8).

7. `quietly generate firm = (uniform() < 0.25) & (x < 5 & y < 5)`
8. `quietly generate nbunits = rpoisson(1) + 1`

Il est possible de cartographier, de manière rudimentaire sur un graphe en deux dimensions (2D), les points simulés ainsi que leurs tailles en termes d'unités de logements et d'unités commerciales (Figure 2).<sup>4</sup>

<sup>2</sup> À noter que l'algorithme peut être généralisé en ajoutant divers critères de centralités ou encore en modifiant la valeur critique des décomptes en fonction de la taille des villes ou municipalités. Les valeurs proposées sont des barèmes qui peuvent être ajustés selon le type de ville/région étudiée : elles peuvent être plus faibles dans des milieux moins urbains, et plus élevées dans des milieux moins ruraux.

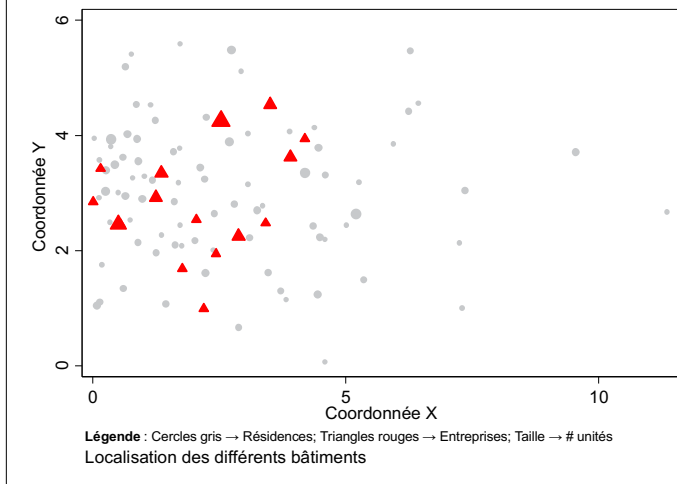
<sup>3</sup> À notre connaissance, ce travail n'a jamais encore été réalisé sur ce logiciel, alors que de nombreuses applications sont développées dans R (Hahsler et al., 2019; Kopczevska et al., 2017).

<sup>4</sup> Cette carte est obtenue avec la commande suivante :

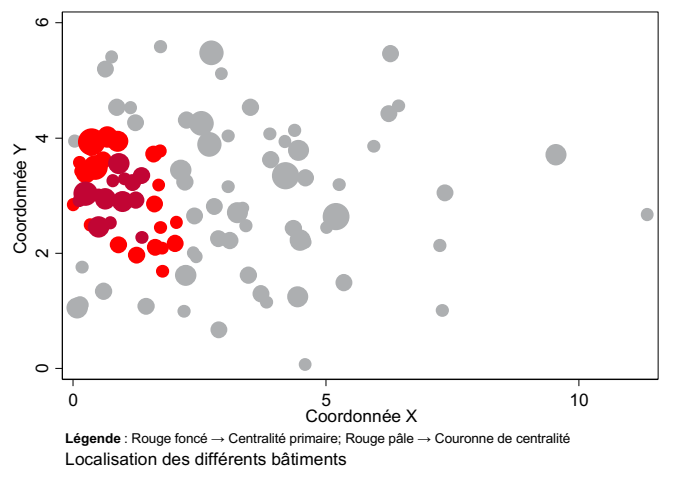
```
#delimit ;
twoway   (scatter y x if firm==0 [fweight = nbunits],
          mcolor(gs12) msize(vsmall) msymbol(circle))
         (scatter y x if firm==1 [fweight = nbunits],
          mcolor(red) msize(small) msymbol(triangle)),
          xtitle(Coordonnée X)
          ytitle(Coordonnée Y)
          legend(off)
          note("bf:Légende) : Cercles gris {&rarr} Résidences; Triangles
              rouges {&rarr} Entreprises; Taille {&rarr} # unités")
          caption("Localisation des différents bâtiments")
          scheme(sicolor) scale(1);
#delimit cr
```

À noter que les options `delimit` permettent d'étaler une commande sur plusieurs lignes. La fin de la commande est identifiée par le symbole `;` (`delimit ;`). La commande `#delimit cr` permet de revenir au format où une ligne représente une commande à exécuter par le logiciel.

**Figure 2.** Localisation des bâtiments (résidences et commerces) fictifs (monocentrique)



**Figure 3.** Localisation des types de centralités économiques (cas monocentrique)



Une fois les données de base simulées et la ville fictive créée, l'algorithme peut être programmé. Pour les fins de l'exercice, seule une forme de centralité (ici primaire) est fixée en fonction des paramètres fondamentaux. La distance critique de recherche est fixée à un rayon de 0,75 unité (ligne 9 -  $d_C = 0,75$ ), le nombre d'unités de logement minimum présent autour est fixé à 15 (ligne 10 -  $\tau_I = 15$ ) et le ratio commerces/logements est fixé à 1 sur 5 (ligne 11 -  $\theta = 1/5$ ). À noter que ces paramètres sont déclarés dans des macros (global) qui permettent de stocker l'information en mémoire et de la rappeler un peu plus loin.

```
**Identifier les paramètres à utiliser**
9. global dcritic = .75
10. global tho = 15
11. global theta = 1/5
```

L'algorithme consiste à calculer le nombre d'unités de logement et d'unités de commerce autour de chacun des points. Ces informations sont enregistrées dans des variables créées au préalable (lignes 12 et 13). Elles sont laissées vides au départ, et seront documentées, à tour de rôle, dans une boucle qui consiste à remplacer les valeurs manquantes par des informations issues de statistiques descriptives.

```
**Créer des variables à meubler
12. quietly generate NbFirms = .
13. quietly generate NbHouse = .
```

La première étape consiste à identifier le nombre de points à documenter (ligne 14) et de conserver ces informations dans des macros (lignes 15 et 16).<sup>5</sup>

```
/*Début de la boucle 1*/
14. summarize id
15. global nmin = r(min)
16. global nmax = r(max)
```

La boucle de calcul s'organise sur l'ensemble des points  $i$  visés (ligne 17) pour prendre fin une fois que l'ensemble des commandes sont exécutées (ligne 29). Pour chacun des points, il s'agit de conserver en mémoire les coordonnées  $x$  (ligne 18) et  $y$  (ligne 20) dans des macros (lignes 19 et 21, respectivement). La distance à l'ensemble des autres points est calculée sur la base de la distance euclidienne (ligne 22).<sup>6</sup> Les données sont ensuite triées de façon croissante sur les distances calculées (ligne 23). En utilisant l'information sur la distance critique ( $d_C$ ), il est possible d'obtenir les informations sur les

statistiques descriptives sur les unités de logement (ligne 24) et sur les unités de commerce (ligne 26). Ces informations sont utilisées afin de documenter les valeurs vides (lignes 25 et 27) des variables créées précédemment, et ce pour chacun des points  $i$ . La variable de distance est ensuite supprimée (ligne 28) pour être recalculée, et ce pour l'ensemble des points  $i$ .

```
**Calcul du nombre de plus proches voisins
17. forvalues i = $nmin/$nmax {
18. quietly summarize x if id==`i'
19. global xi = r(mean)
20. quietly summarize y if id==`i'
21. global yi = r(mean)
22. quietly generate distance = sqrt((x - $xi)^2 + (y - $yi)^2)
23. sort distance
24. quietly summarize nbunits if distance <= $dcritic & firm==1
25. quietly replace NbFirms = r(sum) if id==`i'
26. quietly summarize nbunits if distance <= $dcritic & firm==0
27. quietly replace NbHouse = r(sum) if id==`i'
28. drop distance
29. }
30. sort id
```

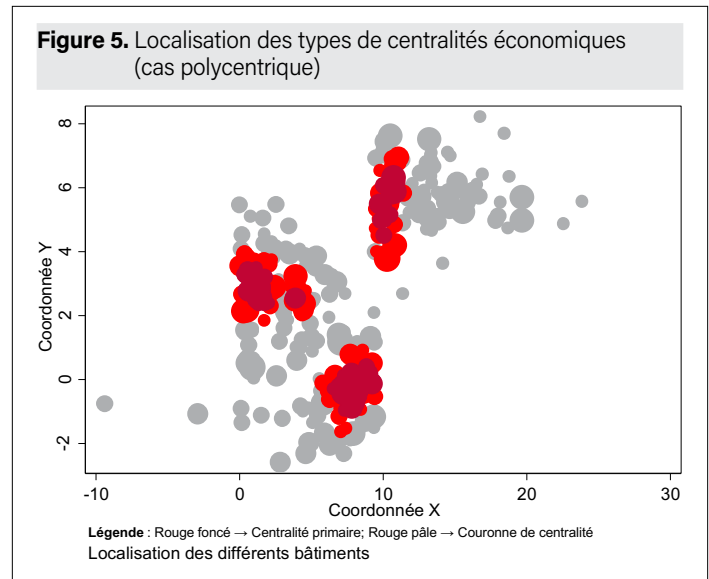
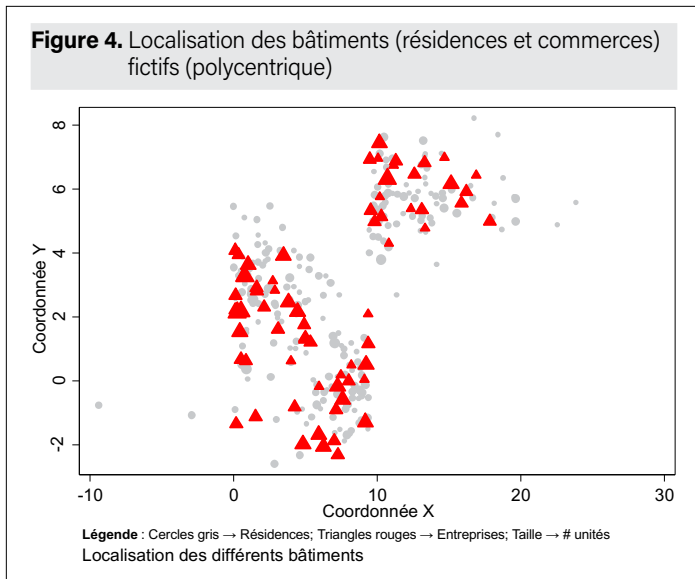
Une fois la boucle complétée, les critères fixés au départ sont appliqués sur les deux variables documentées (ligne 31). Les critères permettent ainsi d'obtenir une nouvelle variable, *center*, qui permet d'identifier les points qui sont considérés comme faisant partie d'une centralité (primaire) économique (oui : 1 / non : 0).

```
31. quietly generate center = (NbHouse)>$tho & NbFirms>($tho*$theta)
```

Une fois les points faisant partie de la centralité identifiés, une seconde boucle est programmée afin d'identifier les points « à proximité » de ces localisations centrales. Puisque la boucle n'a pas besoin de couvrir les points faisant partie de la centralité, il est possible de diminuer le nombre de répétitions à effectuer dans la boucle en créant une variable temporaire. En triant les observations selon leur statut de centralité (ligne 32), on peut isoler les points qui ne sont pas qualifiés de centraux (lignes 33 et 34), et conserver ces informations en mémoire dans deux macros (lignes 35 et 36).

<sup>5</sup> À noter que les macros sont appelées en utilisant le symbole « \$ » devant le nom de la macro, comme c'est le cas pour l'identification des bornes sur lesquelles appliquer la boucle (ligne 17).

<sup>6</sup> Cette distance fait référence au théorème de Pythagore.



```
/*Début de la boucle 2*/
32. sort center
33. quietly generate idtemp = _n
34. summarize idtemp if center==0
35. global mmin = r(min)
36. global mmax = r(max)
```

Comme précédemment, la seconde boucle nécessite la création d'une nouvelle variable (ligne 37) qui permet d'identifier les points situés à moins de la distance critique identifiée ( $dC$ ) des points de centralité. Pour chacun des points qui ne sont pas qualifiés de centralités (ligne 38), la boucle conserve en mémoire les coordonnées géographiques des points (lignes 39 à 42) afin de calculer la distance à l'ensemble des autres points (ligne 43). La distance seuil permet de vérifier si un point faisant partie d'une centralité se trouve dans la zone critique fixée (ligne 45). Si tel est le cas, alors le point est qualifié de couronne de centralité (ligne 46). Autrement, les points seront qualifiés de non centraux de manière résiduelle (ligne 49).

```
**Calcul du nombre de plus proches voisins
37. quietly generate extension = .
38. forvalues j = $mmin/$mmax {
39. quietly summarize x if idtemp==`j'
40. global xj = r(mean)
41. quietly summarize y if idtemp==`j'
42. global yj = r(mean)
43. quietly generate distance = sqrt((x - $xj)^2 + (y - $yj)^2)
44. sort distance
45. quietly summarize center if distance <= $dcritic
46. quietly replace extension = 1 if r(sum)!=0 & idtemp==`j'
47. drop distance
48. }
49. quietly replace extension = 0 if extension==.
```

Une fois les deux boucles complétées, on peut afficher la carte des centralités résultante (Figure 3). La forme des types de centralité n'est pas nécessairement concentrique même si la distance est issue d'une recherche circulaire. Ceci confère d'ailleurs un avantage à l'algorithme: il est capable de détecter plusieurs formes géométriques de centralités.

### Le cas polycentrique

Tout comme précédemment, la première étape est de localiser les points qui seront mobilisés dans l'analyse. Puisque l'exercice repose sur une ville polycentrique, la taille de l'échantillon est augmentée (ligne 50).

```
50. set obs 300
51. quietly generate id = _n
```

Les coordonnées sont ensuite simulées de manière à faire ressortir trois centralités (lignes 52 et 53; lignes 58 et 59; lignes 60 et 61). Les caractéristiques des points sont simulées une fois les localisations établies (lignes 62 et 63).

```
52. quietly generate x = rexponential(3)
53. quietly generate y = rnormal(3)
54. quietly summarize x
55. global tempx = r(max)
56. quietly summarize y
57. global tempy = r(max)
58. quietly replace x = ($tempx/2) + x if id>(_N/3) & id<(_N*(2/3))
59. quietly replace y = ($tempy/2) + y if id>(_N/3) & id<(_N*(2/3))
60. quietly replace x = ($tempx/2) - x if id>=(_N*(2/3))
61. quietly replace y = ($tempy/2) - y if id>=(_N*(2/3))
62. quietly generate firm = (uniform())<0.25)
63. quietly generate nbunits = rpoisson(1) + 1
```

On peut ensuite cartographier la distribution spatiale des points en fonction de leurs caractéristiques respectives (Figure 4).

Une fois les localisations et caractéristiques simulées, il s'agit d'appliquer le code permettant de faire tourner les deux boucles (le code est identique – voir lignes de commande 9 à 49). Le résultat retourne les différents types de centralités de la ville polycentrique fictive. Trois centralités primaires émergent, chacune ayant une couronne de centralité qui lui est propre (Figure 5).

Voyons maintenant comment cet algorithme permet d'identifier les centralités pour différents secteurs de la province de Québec.



## UNE APPLICATION POUR LES AGGLOMÉRATIONS URBAINES DU QUÉBEC

Une analyse est menée sur les six (6) régions métropolitaines de recensement (RMR) ainsi que les vingt-cinq (25) agglomérations de recensement (AR) de la province. Pour chacune des régions identifiées, les informations provenant des rôles d'évaluation, fournis par le Ministère des Affaires municipales et de l'habitation (MAMH) pour l'année 2020 sont mobilisées. Les rôles d'évaluation municipaux permettent de tirer profit des informations portant sur les unités d'évaluation individuelles.

**Tableau 2.** Liste des RMR et AR retenues

Nom des aires sélectionnées	# municipalités	# unités d'évaluation retenues	Temps de calcul (heures)†
<b>Région Métropolitaine de recensement (RMR)</b>			
Gatineau	12	124 706	58,2095
Montréal	87	1 241 418	> 2 semaines
Québec	26	267 555	> 4 jours
Saguenay	8	61 066	13,5392
Sherbrooke	11	76 703	20,7231
Trois-Rivières	6	57 381	11,4192
<b>Agglomérations de recensement (AR)</b>			
Alma	2	12 360	0,4261
Amos	6	6 796	0,1414
Baie-Comeau	6	11 419	0,4072
Campbelton	1	703	0,0019
Cowansville	1	4 367	0,0542
Dolbeau-Mistassini	3	6 719	0,1328
Drummondville	8	33 257	3,5553
Granby	4	29 356	2,7297
Hawkesbury	1	707	0,0017
Joliette	4	16 316	0,7328
Lachute	1	4 827	0,0686
Matane	4	8 612	0,2358
Rimouski	3	20 722	1,3800
Rivière-du-Loup	5	11 235	0,3742
Rouyn-Noranda	2	18 615	1,1203
Sallaberry-de-Valleyfield	1	14 956	0,6661
Sept-Îles	1	9 726	0,2814
Shawinigan	3	21 722	1,4319
Sorel-Tracy	4	18 939	1,0333
Saint-Georges	5	15 202	0,6836
Saint-Hyacinthe	3	19 481	1,0975
Saint-Jean-sur-le-Richelieu	1	31 646	3,0819
Thetford Mines	3	12 571	0,4431
Val-d'Or	1	12 797	0,4886
<b>Total</b>	<b>223</b>	<b>2 171 880</b>	<b>124,46</b>

† : Ordinateur utilisé Intel® Core™ i7-6700 CPU @3,40 GHz 48 Go RAM

7 Correspondant au CUBF supérieur à 9000

8 Correspondant aux CUBF entre 4500 et 4600.

9 Ce rayon est obtenu par le calcul de l'aire d'un cercle (aire =  $\pi r^2$ , où  $\pi$  est une constante et  $r$  est la longueur du rayon).

10 Afin de faciliter la présentation, seules les cartes pour les six RMR sont présentées dans la note. Les autres cartes sont disponibles en annexe et sur demande auprès de l'auteur correspondant.

Les rôles d'évaluations à l'échelle de la province contiennent plus de 3,6 millions d'unités d'évaluation. Afin de faciliter et d'accélérer les calculs, les unités sont d'abord isolées pour chacune des 31 régions identifiées (Tableau 2). Au total, les 31 régions retenues couvrent 233 municipalités, ce qui représente un peu plus de 2 millions d'unités d'évaluation, soit pratiquement le tiers des unités totales de la province. Afin d'éliminer des observations inutiles dans les calculs, les unités d'évaluation non construites (ou vacantes<sup>7</sup>) et les unités d'évaluation identifiant les infrastructures routières<sup>8</sup> sont éliminées.

L'algorithme est appliqué de manière indépendante sur chacune des régions. À noter que l'algorithme pour la RMR de Montréal est scindé en quatre composantes: l'île de Montréal, l'île de Laval, la Rive-Sud et la Rive-Nord. Cette approche permet non seulement d'accélérer les calculs, mais également d'éviter de prendre des unités de part et d'autre du fleuve pour former une centralité et ainsi respecter les contraintes naturelles de la RMR.

L'algorithme est divisé en deux boucles. La première permet d'identifier les centralités primaires ( $\tau_1$  et  $\pi_1$ ) et secondaires ( $\tau_2$  et  $\pi_2$ ) sur la base des critères préétablis pour un cercle d'un kilomètre carré ( $d_C = 565$  mètres).<sup>9</sup> La seconde permet d'isoler, pour les unités d'évaluation qui ne sont pas qualifiées de centralités, celles qui se trouvent à proximité des unités qualifiées de centralités. Une fois les deux boucles effectuées, les unités résiduelles, c'est-à-dire qui ne font ni partie d'une centralité (primaire et secondaire) ni d'une couronne de centralité, forment le quatrième type: les unités hors centralités.

L'exécution de l'algorithme se fait assez rapidement pour des RMR/AR de taille modérée. Par exemple, pour Rimouski, le temps d'exécution pour l'ensemble des unités d'évaluation est d'une heure 22 minutes (Tableau 2). En revanche, le temps de calcul est considérable lorsque le nombre d'unités augmente. Le temps nécessaire pour le calcul des deux boucles pour Gatineau est pratiquement de 2 jours et demi (Tableau 2).

L'application de l'algorithme permet d'obtenir un ensemble de cartes permettant d'identifier les points (unités d'évaluation) qui font partie des différents types de centralité. Il permet d'avoir un portrait qui est essentiellement descriptif, mais qui donne un portrait global des centralités dans les différentes régions (Figures 6)<sup>10</sup>.

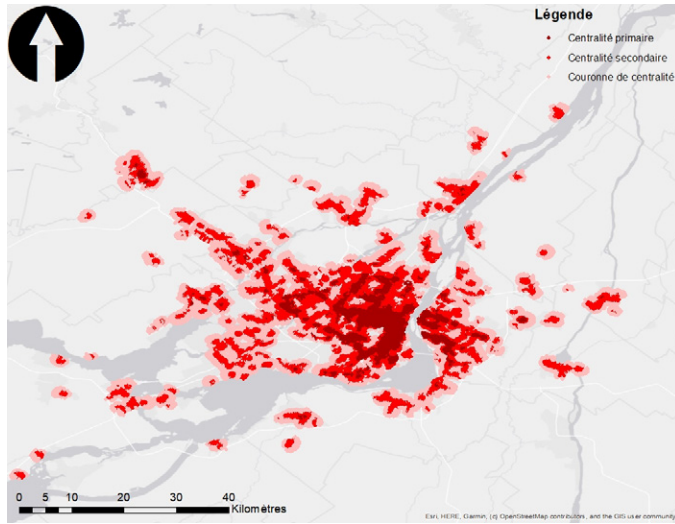
Les cartes mettent en relief l'historique qui a façonné les villes et les centralités à l'heure actuelle. En ce sens, elles peuvent être utilisées afin de décrire les schémas de localisation qui ont mené à la structure actuelle. Elles peuvent aussi être mobilisées afin d'identifier les concentrations et les axes qui structurent les dynamiques locales actuelles. Elles permettent d'identifier des corridors de développement, mais aussi des points de rupture. Elles permettent également d'identifier les RMR/AR qui ont encore une structure monocentrique, où celles qui s'en éloignent lentement (ou récemment).

Elles peuvent également servir à identifier la localisation des différents centres afin d'exprimer une variable indépendante dans un modèle statistique (la distance aux différents centres). Il est ainsi possible d'inclure les principaux pôles susceptibles de générer et d'attirer de nombreux déplacements dans la ville.

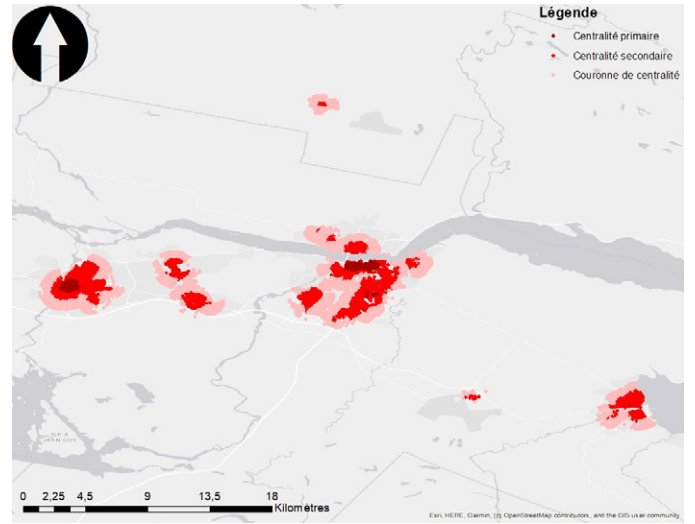
Les statistiques descriptives identifiant les unités d'évaluation classées dans chacun des types de centralités (Tableau 3), permettent aussi un exercice intéressant. Elles servent de référence afin de comparer les RMR et RA entre elles. Elles permettent d'identifier des RMR ou des RA plus denses, en relativisant le nombre d'unités d'évaluation classé comme centralité primaire et/ou secondaire par rapport aux unités résiduelles.



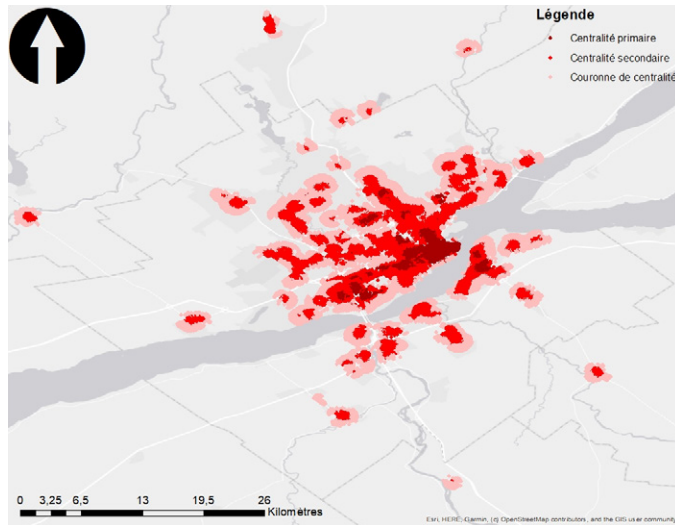
**Figure 6.** LCentralités détectées dans les RMR du Québec



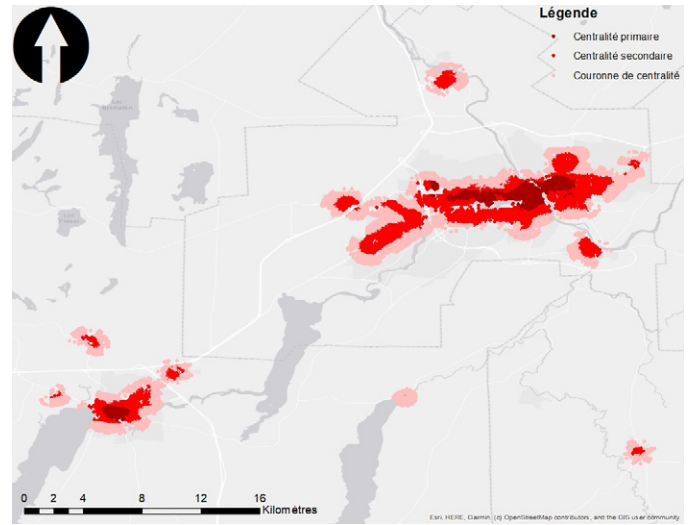
Montréal



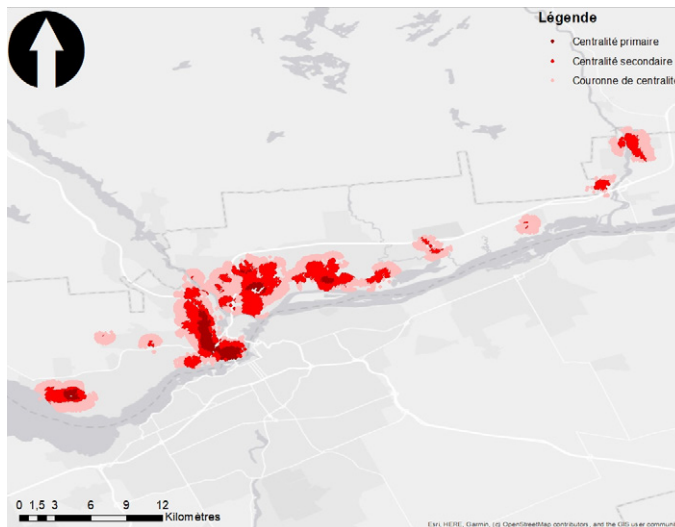
Saguenay



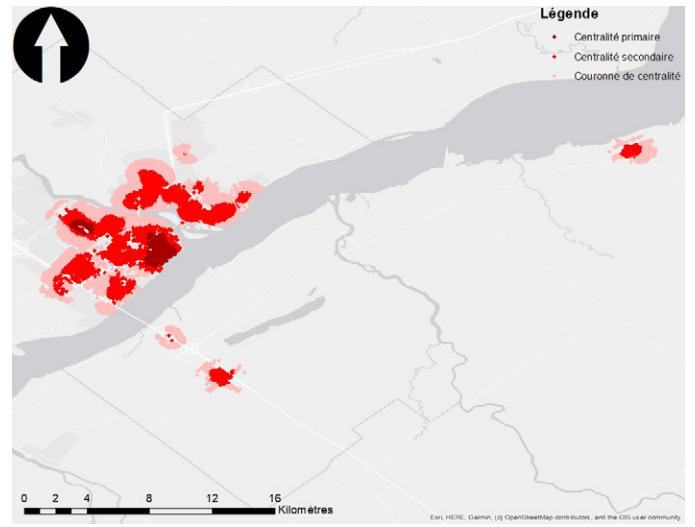
Québec



Sherbrooke



Gatineau



Trois-Rivières

**Tableau 3.** Nombre d'unités d'évaluation par type de centralités et RMR/AR

Nom des aires sélectionnées	# unités de centralité primaire	# unités de centralité secondaire	# unités de couronne de centralités	# unités résiduelles
<b>Région Métropolitaine de recensement (RMR)</b>				
Gatineau	5 124	18 907	24 259	76 416
Montréal	267 971	391 810	321 250	260 387
Québec	23 501	66 503	74 319	103 232
Saguenay	1 340	9 659	14 818	35 249
Sherbrooke	5 203	16 370	19 241	35 889
Trois-Rivières	2 980	12 964	11 715	29 722
<b>Agglomérations de recensement (AR)</b>				
Alma	600	3 407	3 131	5 222
Amos	0	1 274	1 453	4 069
Baie-Comeau	91	1 902	3 404	6 022
Campbelton	0	0	0	703
Cowansville	149	1 394	1 688	1 136
Dolbeau-Mistasini	0	1 535	1 566	3 618
Drummondville	1 604	8 113	9 752	13 788
Granby	1 824	6 088	7 086	14 358
Hawkesbury	0	57	436	214
Joliette	823	4 986	4 797	5 710
Lachute	253	982	1 696	1 896
Matane	394	1 453	1 383	5 382
Rimouski	1 516	2 320	3 695	13 191
Rivière-du-Loup	351	2 420	2 114	6 350
Rouyn-Noranda	1 556	1 872	1 802	13 385
Sallaberry-de-Valleyfield	1 654	3 343	3 793	6 166
Sept-Îles	430	2 095	1 466	5 735
Shawinigan	665	5 468	4 809	10 780
Sorel-Tracy	814	4 529	5 913	7 683
Saint-Georges	530	3 848	3 871	6 953
Saint-Hyacinthe	1 536	4 929	5 953	7 063
Saint-Jean-sur-le-Richelieu	1 255	9 460	9 417	11 514
Thetford Mines	46	4 420	2 866	5 239
Val-d'Or	1 096	1 949	2 118	7 634
<b>Total</b>	<b>323 306</b>	<b>594 057</b>	<b>549 811</b>	<b>704 706</b>

Bref, bien que l'algorithme permette de produire des informations et des cartes qui sont essentiellement descriptives, les résultats peuvent être mobilisés afin de créer des informations utiles dans des analyses statistiques.

## CONCLUSION

La présente note avait pour but de présenter un algorithme permettant de localiser les centralités (économiques) dans différentes villes et agglomérations de la province à partir des données des rôles d'évaluation. Pour ce faire, une approche inspirée de la méthode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) est proposée en utilisant des micro-données. L'algorithme propose une typologie des centralités en quatre catégories : les centralités primaires, les centralités secondaires, les couronnes de centralités, et l'absence de centralité (les résidus).

La mise en place de l'algorithme est d'abord proposée en utilisant des données fictives. Deux jeux de données sont développés : i) un premier pour une ville monocentrique; et ii) un second pour une ville polycentrique. L'algorithme est présenté et expliqué étape par étape pour ensuite être appliqué sur les régions métropolitaines de recensement (RMR) et les agglomérations de recensement (AR) de la province en utilisant les données issues des rôles d'évaluations. Les résultats permettent une cartographie complète des centralités économiques et identifient, par le fait même, les secteurs les plus denses en termes d'unités d'habitations et d'unités commerciales.

L'originalité du travail est d'opérationnaliser une approche qui permet d'identifier des centralités économiques en recourant aux micro-données spatiales avec l'aide de quelques paramètres clés : le nombre d'unités de logement et le nombre de commerce. L'algorithme ne nécessite aucun a priori particulier sur la forme des polygones, ou encore sur la localisation du centre historique. Le fait d'utiliser des micro-données spatiales permet de généraliser les approches classiques qui prennent habituellement appui sur des données spatiales agrégées (polygones) afin d'en tirer un portrait d'une plus grande précision.

Pour le moment, l'approche présentée se concentre uniquement sur deux variables. Elle peut néanmoins être étendue afin d'ajouter des critères supplémentaires. Il est possible d'ajouter des informations sur la hauteur des bâtiments, ou encore sur la présence d'infrastructures particulières, sur la présence d'institutions publiques, voire politiques, et sur l'attractivité des lieux. Sans être exhaustifs, ces ajouts permettraient de raffiner l'analyse et d'enrichir la typologie des centralités qui, pour le moment, reflète exclusivement les centralités économiques.

Néanmoins, certaines faiblesses demeurent. D'une part, bien que l'application soit simple, le temps de calcul peut être important, surtout pour les RMR de grandes tailles, ce qui en constitue une des plus grandes faiblesses. Cette faiblesse peut néanmoins être contournée en utilisant une approche par échantillonnage (Arribas-Bel et al., 2021). D'autre part, les données portant sur le nombre d'unités commerciales contenues dans les unités d'évaluation peuvent sous-estimer l'importance des activités économiques ayant lieu sur place. C'est le cas pour les activités économiques qui sont de grandes consommatrices d'espace (manufactures et commerce de gros). Une application propre à la détection de regroupements (*clusters*) pour ce type d'entreprises s'avère une approche intéressante afin de bonifier l'algorithme existant.

Elles permettent de juger de la densité de la centralité primaire, en la comparant au nombre d'unités d'évaluation classées comme centralité secondaire. Elles permettent également de juger de la concentration spatiale de la centralité, en prenant le ratio des différents types de centralité par rapport aux unités d'évaluation résiduelles.

Appliqué à des années différentes, l'algorithme permet de juger, dans un contexte spatio-temporel, de l'évolution des centralités au fil du temps. Cet exercice permet, notamment, de documenter le phénomène d'étalement urbain. Il peut également servir à comparer la localisation et la taille des différents types de centralité au fil des années afin de jauger la croissance des agglomérations sur les critères de la concentration de l'écoumène.

## RÉFÉRENCE

- Allain, R. (2004). *Morphologie urbaine. Géographie, aménagement et architecture de la ville*, Armand Colin, Paris.
- Alonso, W. (1964). *Location and Land Use, Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, 205 p.
- Anselin, L. (1995). Local Indicators of Spatial Association – LISA, *Geographical Analysis*, 27(2): 93-115.
- Arribas-Bel, D., Garcia-Lopez, M.-A. & Viladecans-Marsal, E. (2019). Building(s and) Cities: Delineating Urban Areas with a Machine Learning Algorithm, *Journal of Urban Economics*, doi: <https://doi.org/10.1016/j.jue.2019.103217>
- Ascher, M. (1999). *Why Governments Waste Natural Resources: Policy Failures in Developing Countries*, John Hopkins University Press, Baltimore.
- Barnes, T. (2001). "In the Beginning was Economic Geography" – A Science Studies Approach to Disciplinary History, *Progress in Human Geography*, 25(4): 521-544.
- Baumont, C., Bourdon, F. & Guillain, R. (2004). Mutations urbaines et logiques de localisation des emplois: le cas de la Communauté de l'agglomération dijonnaise (1990, 1999), *Revue d'économie régionale et urbaine*, 4: 579-608.
- Baumont, C., Combes, P.-P., Derycke, P.-H. & Jayet, H. (2000). *Économie géographique: Les théories à l'épreuve des faits*, Economica, Paris, 326 p.
- Beaujeu-Garnier, J. (1967). Méthode d'étude pour le centre des villes, *Annales de géographie*, 74(406): 695-707.
- Bingham, R.D. & Kimble, D. (1995). The Industrial Composition of Edge Cities and Downtowns: The New Urban Reality, *Economic Development Quarterly*, 9(3): 259-272.
- Bogart, W. & Ferry, W. (1999). Employment Centres in Greater Cleveland: Evidence of Evolution of a Formerly Monocentric City, *Urban Studies*, 36: 2099-2110.
- Bourdeau-Lepage, L., Hurriot, J.-M. & Perreur, J. (2009). À la recherche de la centralité perdue, *Revue d'économie régionale et urbaine*, 3: 549-572.
- Castells, M. (1973). *La question urbaine*, François Maspero, Paris.
- Cervero, R. & Wu, K. (1997). Polycentrism, Commuting, and Residential Location in the San Francisco Bay Area, *Environment and Planning A*, 29: 865-886.
- Cressie, N. (1991). *Statistics for Spatial Data*, Wiley Classic Library, 928 p.
- Chalas, Y. (2010). La ville de demain sera une ville-nature, *L'Observatoire*, 37(2): 3-10.
- Claval, P. (1981). *La logique des villes. Essai d'urbanologie*, Litec, Paris, 633 p.
- Craig, S. & Ng, P. (2001). Using Quantile Smoothing Splines to Identify Employment Subcenters in a Multicentric Urban Area, *Journal of Urban Economics*, 49: 100-120.
- Devisme, L. (2005). *La ville décentrée. Les figures centrales à l'épreuve des dynamiques urbaine*, L'Harmattan, Paris.
- De Bellfon, M.-P., Combes, P.-P., Duranton, G., Gobillon, L. & Gorin, C. (2019). Delineating Urban Areas using Building Density, *Journal of Urban Economics*, <https://doi.org/10.1016/j.jue.2019.103226>
- Dubé, J. & Legros, D. (2014). *Introduction à l'économétrie spatiale des micro-données*, Éditions Hermes Penton Ltd (ISTE), Londres, 228 p.
- Dubé, J. & Brunelle, C. (2014) Dots to Dots: A General Methodology to Build Local Indicators using Spatial Micro Data, *Annals of Regional Science*, 53(1): 245-272.
- Duranton, G. (1997). La nouvelle économie géographique: agglomération et dispersion, *Économie & Prévision*, 131(5): 1-24.
- Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996). Density-based Spatial Clustering of Applications with Noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 240: 226-231.
- Fujita, M. & Thisse, J.-F. (1997). Économie géographique, Problèmes anciens et nouvelles perspectives, *Annales d'économie et de statistique*, 45: 37-87.
- Galan, S.F. (2019). Comparative Evaluation of Region Query Strategies for DBSCAN Clustering, *Information Sciences*, 502: 76-90.
- Gaschet, F. et Lacour, C. (2002). Métropolisation, centre et centralité, *Revue d'économie régionale et urbaine*, 1: 49-72.
- Gauthiez, B. (2004). The History of Urban Morphology, *Urban Morphology*, 8(2): 71-89.
- Gbenyo, K. & Dubé, J. (2018). Le MAUP et l'insoluble question de frontière optimale, *Revue d'Économie Régionale et Urbaine*, 3: 619-649.
- Gehl, J. (2012). *Pour des villes à échelle humaine*, Écosociété, 212 p.
- Giuliano, R., Redfern, C.L., Agarwal, A., Li, C. & Zhuang, D. (2007). Employment Concentrations in Los Angeles, 1980-2000, *Environment and Planning A*, 39(12): 2935-2957.
- Giuliano, G. & Small, K. (1991). Subcenters in the Los Angeles Region, *Regional Science & Urban Economics*, 21(2): 163-182.
- Griffith, D.A. (1981). Modelling Urban Population Density in a Multi-centred City, *Journal of Urban Economics*, 9: 298-310.
- Griffith, D.A. & Wong, D. (2007). Modeling Population Density across Major US Cities: A Polycentric Spatial Regression Approach, *Journal of Geographical Systems*, 9(1): 53-75.
- Grafmeyer, Y. & Joseph, I. (2004). *L'École de Chicago: naissance de l'écologie urbaine*, Champ Urbain.
- Guillain, R. & Le Gallo, J. (2010). Agglomeration and Dispersion of Economic Activities in and Around Paris: An Exploratory Spatial Data Analysis, *Environment and Planning B*, 37(6): 961-981.
- Guillain, R., Le Gallo, J. & Boiteux-Orain, C. (2006). Changes in Spatial and Sectoral Patterns of Employment in Ile-de-France, 1978-97, *Urban Studies*, 43(11): 2075-2098.
- Hahsler, M., Piekenbrock, M. & Doran, D. (2019). DBSCAN: Fast density-based Clustering with R, *Journal of Statistical Software*, 91(1): 1-30.
- Halbert, L. (2006). The Polycentric City Region that Never was: The Paris Agglomeration, Bassin Parisien and Spatial Planning Strategies in France, *Built Environment*, 32(2): 183-193.
- Hoover, E. (1948). *The Location of Economic Activity*, McGraw Hill, New York.
- Hotelling, H. (1929). Stability in Competition, *Economic Journal*, 39: 41-57.
- Khan, K., Rehman, S.U., Aziz, K., Fong, S. & Sarasvady, S. (2014). DBSCAN: Part, Present and Future, *Proceedings of the International Conference on the Applications of Digital Information and Web Technologies*, IEEE, 232-238.
- Kopczewska, K. (2018). Cluster-based Measures of Regional Concentration. Critical Overview, *Spatial Statistics*, 27: 31-57.



- Kopczewska, K., Churski, P., Ochojski, A. & Polko, A. (2017). *Measuring Regional Specialisation – A New Approach*, Palgrave Macmillan, London.
- Kulldorf, M. (1997). A Spatial Scan Statistics, *Communications in Statistics: Theory and Methods*, 26: 1481-1496.
- Lebrun, N. (2004). *Centralités urbaines et concentrations de commerces*, Thèse de doctorat, Université de Reims, Reims.
- Lévy, J. (1994). *L'espace légitime. Sur la dimension géographique de la fonction politique*, Presses de la fondation nationale des sciences politique, Paris.
- McDonald, J.F. (1997). *The Fundamentals of Urban Economics*, Prentice Hall, 545 p.
- McDonald, J.F. (1987). The Identification of Urban Employment Subcenters, *Journal of Urban Economics*, 9: 298-310.
- McDonald, J.F. & Prather, P.J. (1994). Suburban Employment Centers: The case of Chicago, *Urban Studies*, 31: 201-218.
- McMillen, D.P. (2003). Identifying Sub-centers using Contiguity Matrices, *Urban Studies*, 40(1): 57-69.
- McMillen, D.P. (2001). Nonparametric Employment Subcenter Identification, *Journal of Urban Economics*, 50(3): 448-473.
- McMillen, D.P. & McDonald, J.F. (1998). Suburban Subcenters and Employment Density in Metropolitan Chicago, *Journal of Urban Economics*, 43: 157-180.
- McMillen, D.P. & McDonald, J.F. (1997). A Non-parametric Analysis of Employment Density in a Polycentric City, *Journal of Regional Science*, 37: 591-612.
- McMillen, D.P. & McDonald, J.F. (1990). Employment Subcenters and Land Values in a Polycentric Urban Area: The case of Chicago, *Environment and Planning A*, 22: 1561-1574.
- McMillen, D.P. & Smith, S.C. (2003). The Number of Subcenters in Large Urban Areas, *Journal of Urban Economics*, 53(3): 321-338.
- Merlin, P. & Choay, F. (2000). *Dictionnaire de l'urbanisme et de l'aménagement*, Presses universitaires de France, Paris, 901 p.
- Murphy, R.R. & Vance, J.E. Jr (1954). Delimiting the CBD, *Economic Geography*, 30(3); 189-222.
- Mills, E. (1967). An Aggregate Model of Resource Allocation in a Metropolitan Area, *American Economic Review*, 57: 197-211.
- Monnet, J. (2000). Les dimensions symboliques de la centralité, *Cahiers de géographie du Québec*, 44(123): 399-418.
- Muth, R.F. (1969). *Cities and Housing*, University of Chicago Press, Chicago, USA.
- OECD (2013). Definition of Functional Urban Areas (FUA) for the OECD Metropolitan Database ([Definition-of-Functional-Urban-Areas-for-the-OECD-metropolitan-database.pdf](#)).
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem: Concepts and Techniques in Modern Geography*, Geo Books.
- Polèse, M., Shearmur, R. & Terral, L. (2015). *Économie urbaine et régionale, Géographie économique et dynamique des territoires*, Economica, Paris, 416 p.
- Redfern, C.L. (2007). The Topography of Metropolitan Employment: Identifying Centers of Employment in a Polycentric Urban Area, *Journal of Urban Economics*, 61(3): 519-541.
- Roncayolo, M. (1990). *La ville et ses territoires*, Folio Essais, Gallimard, Paris.
- Rozenfeld, H.D., Rybski, D., Gabaix, X. & Makse, H.A. (2011). The Area and Population of Cities: New Insights from a Different Perspective on Cities, *American Economic Review*, 101(5): 2205-2225.
- Sassen, S. (1991). *The Global City*, Princeton University Press, Princeton.
- Simard, M. (2012). Urbain, rural et milieux transitionnels: les catégories géographiques de la ville diffuse, *Cahiers de géographie du Québec*, 56(157): 109-124.
- Small, K. & Song, S. (1994). Population and Employment Densities: Structure and Change, *Journal of Urban Economics*, 36: 292-313.
- Statistique Canada (2016). Classification des centres de population et des régions rurales 2016, en ligne ([Population Centre and Rural Area Classification 2016 \(statcan.gc.ca\)](#)).
- Storper, M. & Venables, A.J. (2004). Buzz: Face-to-face Contact and the Urban Economy, *Journal of Economic Geography*, 4(4): 351-370.
- Vant, A. (1971). L'objectif et le subjectif: problème de délimitation du centre de Saint-Étienne, *Revue de géographie de Lyon*, 46(2): 199-225.
- Weber, A., (1929). *Theory of the Location of Industry*. Chicago University Press, Chicago.