Université catholique de Louvain
Faculté des Sciences, École de Géographie
Center for Operations Research and Econometrics

Exploring new geographies of interactions in and around the metropolitan area of Brussels

---

Doctoral Dissertation presented by

Arnaud Adam

in fulfillment of the requirements
for the degree of Doctor in Sciences

Thesis committee:

Prof.	Isabelle Thomas (Supervisor)	Université catholique de Louvain
Prof.	Philippe Chevalier (Chair)	Université catholique de Louvain
Prof.	Jean-Charles Delvenne (Secretary)	Université catholique de Louvain
Prof.	Ann Verhestsel	Universiteit Antwerpen
Dr.	Clémentine Cottineau	UMR 8097 Centre Maurice Halbwachs
Dr.	Christophe Cloquet	Poppy
M.	Xavier Dehaibe	Perspective Brussels

Louvain-la-Neuve, Belgium
August 2019

Acknowledgements

It is not every day that we can congratulate ourselves for concluding a work lasting more than four years. Four years during which my research was organised into networks, composed of places and weighted exchanges. However, these networks would have been nothing without the material support of Innoviris. I would also like to thank UCLouvain and especially CORE for the most practical aspect of being a *professional student*.

For her accompaniment during these four years, and for having trusted me by committing myself to the Brunet research contract, I would like to fully thank my promoter, Prof. Isabelle Thomas. Her advice, discussions, questioning, support and availability were key factors in the completion of this thesis. In addition to initiating my career, these years have allowed me to achieve things that I thought impossible, that changed me and that will remain long in my memory.

I also want to thank the other members of my supervisory panel for their availability, their professionalism, and the exchanges we had. I am particularly thinking of Prof. Aan Verhetsel, Prof. Jean-Charles Delvenne and Christophe Cloquet. I would also like to thank the two more recent members of my thesis committee: Clementine Cottineau and Xavier Dehaibe. I can imagine how difficult it must have been to soak up a long-running project. To this thesis, you have brought a well-needed external perspective. I want to thank you both for your availability and the interest you have shown in my project. In general, I would like to thank the members of my thesis committee and its President Prof. Philippe Chevalier: your feedback has greatly helped to improve this manuscript.

During these four years, I was lucky enough to collaborate with the IWPES. I would like to thank them for their warmly welcome and their interest to my work, ending up with a nice and friendly collaboration.

I would also like to thank the COST TU1305 program "Social Network and Travel Behaviour" which have allowed me to exchange and collaborate with European experts from various disciplines. I will not forget my first conference in Budapest. The Turkish and steam baths, as well as sunrise were just the beginning of other European stays, ranging from Bucharest and Bari, to York via Tartu. However, these foreign stays are nothing without the people with whom they are shared. The peak was located in New Orleans, where, between crocodiles, storm, aquariums, and the "keep going" song, I discovered the jazzy life of the United States along with wonderful colleagues: Sonia Trabelsi, Madeleine Guyot, Justin Delloye, Gaetan Montero and Olivier Finance. I do not forget either other *corrist* colleagues such as Adeline Decuyper and Jonathan Jones. Without you, my breaks and geo-en-core meetings would not have been the same. Altogether, we were a very nice group, both around our work/terrace tables, feet in the sand. I would have a stronger thought for Justin Delloye, who besides forcing me to embark on this project, gives me a beautiful friendship.

I do not want to finish my thanks without a thought turned towards a more northern country. I especially want to thank Anniki Puura who, through our discussions, her listening and her presence added a wonderful unexpected dimension to this thesis. I also thank her for the open mind she has instilled in me, our geographical discoveries, the learning of the sauna and the wonderful moments. Before coming back to supplementary Belgian thanks, I would also have a final Estonian thought for Prof. Rein Ahas who unfortunately left us too soon. His humour, quick-wit and intelligence will be missed.

J'aurais finalement une pensée plus générale pour toutes les personnes qui ont traversé ma vie ces dernières années, qu'ils n'aient été que de belles parenthèses ou qu'ils aient toujours été dans mon quotidien. Je ne nommerais personne en particulier par manque de place, mais je vous suis reconnaissant de votre écoute, de votre soutien et de votre compréhension mise à rude épreuve ces derniers mois.

Finalement, je conclurais en remerciant encore plus fort mes parents et ma soeur pour leurs écoutes et présences quotidienne. Votre aide m'a été des plus importantes pour me faire conserver mon courage dans les moments difficiles de ces dernières années, me poussant encore plus à finaliser ce projet.

Thank you, aitäh, merci.

Contents

I Introduction **1**

1 A general background **3**
1.1 Studied area: Belgium and Brussels 4
1.2 Movements, interactions in space 7
1.2.1 Transport geography theory 7
1.2.2 Spatial interaction models 8
1.3 Big-data 10
1.4 Community detection methods 11
1.5 Urban models 12
1.6 Objectives of this contribution 14

II Methodological framework **19**

2 The multi-scale Louvain Method to detect communities: robustness tests **21**
2.1 Introduction 22
2.2 Methodology and data 24
2.2.1 Community detection 24
2.2.2 Data 27
2.2.3 Sensitivity analyses 29
2.3 Results 30
2.3.1 Effect of parameter ρ 31
2.3.2 Effect of self-loops 38
2.3.3 Delineation of the study area 39
2.3.4 Infomap 40
2.4 Conclusion 42

3 Residential basins in Belgium: two methods, one reality? 47
3.1 Introduction . . . 48
3.2 The Anabel method . . . 49
3.3 Application of the two methods to a theoretical network . . . 50
3.3.1 The synthetic network . . . 50
3.3.2 Application of Anabel . . . 50
3.3.3 Application of the Louvain Method . . . 52
3.3.4 Method comparaisons . . . 53
3.4 Application of the two methods to Belgium . . . 56
3.4.1 Data matrix . . . 56
3.4.2 Anabel method Results . . . 56
3.4.3 Louvain Method results . . . 58
3.4.4 Comparison of the results . . . 60
3.5 Conclusion . . . 62

III Movements of people 65

4 Exploring job and residential basins in and around Brussels 67
4.1 Introduction . . . 68
4.2 Data and method . . . 70
4.2.1 Data: migration and commuting . . . 70
4.2.2 Method . . . 71
4.2.3 Defining the extension of Brussels . . . 72
4.3 Results: spatial partitions . . . 72
4.3.1 The case of Belgium . . . 72
4.3.2 The case of Brussels . . . 82
4.4 Conclusion . . . 87

5 Train travel requests: does geography still matter? 91
5.1 Introduction . . . 92
5.2 Data and methodology . . . 93
5.2.1 Urban and railway structures of Belgium . . . 93
5.2.2 Travel demand: the *iRail* data set . . . 94
5.2.3 Methodology . . . 97
5.3 Results . . . 102
5.3.1 All requests . . . 102
5.3.2 Temporal variations . . . 103
5.3.3 Linguistic variations . . . 104
5.3.4 Regional variations: the RER zone . . . 105
5.4 Discussion . . . 107
5.4.1 Urban or network oriented communities? . . . 107
5.4.2 Implications for policy decisions . . . 110

5.4.3 Challenges and paths for future research 111
5.5 Conclusions 112

IV Movements of freight 115

6 Exploring freight transportation through the real-time tracking of trucks 117
6.1 Introduction 118
6.2 From raw GPS data to a spatial information 120
6.2.1 Tracking trucks in space and time 120
6.2.2 From GPS trace to a succession of trips and segments . 122
6.2.3 Delineation of the study area 128
6.3 Brussels within the network of trucks movements 129
6.3.1 Circulation of trucks at the scale of Belgium and Brussels 129
6.3.2 Revealing the economic polarisation of Brussels by the circulation of trucks 132
6.4 Brussels within the (inter)national circulation of trucks 136
6.4.1 Using different networks to catch different realities . . . 136
6.4.2 Determining communities of movements 138
6.5 Conclusion 142

V Movements of information 145

7 Brussels and its periphery: reviving the Chicago School with mobile phone calls? 147
7.1 Introduction 148
7.2 Studied area: Brussels and its surroundings 149
7.3 Clustering places with mobile phone data 149
7.3.1 Mobile phone data 149
7.3.2 Clustering places with tight interactions 152
7.3.3 Clustering places with similar hourly variation 157
7.4 Clustering places with similar characteristics 161
7.5 Combining partitions 164
7.5.1 Methodological issues 165
7.5.2 A concluding map 166
7.6 Conclusion 168

VI Concluding remarks **175**

8 Conclusion **177**
8.1 Summary of findings 178
8.1.1 General objectives of this thesis 178
8.1.2 Data and methodological framework 178
8.1.3 Key results of this thesis 179
8.2 Critical assessment 184
8.3 Paths for further research 186
8.3.1 Linked with community detection method 186
8.3.2 Linked with mobile phone calls 188
8.3.3 Linked with the real-time following of trucks 188
8.4 Implications and recommendations 188
8.5 Concluding words 192

Bibliography **193**

List of Figures **217**

List of Tables **223**

Main Acronyms used in this contribution:

And	Anderlecht
Aud	Auderghem
B	City of Brussels (municipality)
BCR	Brussels Capital Region
BSU	Basic Spatial Unit
ET	Etterbeek
EVE	Evere
For	Forest
ICT	Information and communication technologies
IX	Ixelles
Mol	Molenbeek-Saint-Jean
NMI	Normalized Mutual Information
OBU	On Board Unit
OD	Origin and Destination
Ottignies	Ottignies Louvain-la-Neuve
RER	Réseau Express Régional
S	Schaerbeek
SG	Saint Gilles
SJ	Saint Josse ten Noode
SNCB	Belgian National Railway Company
U	Uccle
VI	Variation of Information
WL	Woluwe-Saint-Lambert
WP	Woluwe-Saint-Pierre
Z	Zaventem

"La ville, c'est un paysage. Au bord d'une rivière que vous ne connaissez pas, sous un climat dont vous ignorez les sautes d'humeur et la forme des nuages. Pourtant c'est de la terre, des arbres, c'est de l'eau, du ciel et du béton, ça devrait se ressembler, les immeubles sont des immeubles et les rues sont des rues, ça ne devrait pas être si différent mais vous ne reconnaissez rien. La hauteur des bâtiments, la largeur des avenues, les logos des marques et les enseignes des boutiques. La couleur des boites aux lettres, des colonnes d'eau, des bouches d'incendies, les uniformes de policiers, des postiers, des livreurs, des caissières, des conducteurs de bus et des ouvriers. La largeur des trottoirs, la place que prend le ciel au-dessus de vous, recouvrant les boulevards et filant dans les allées. La présence des autoroutes en pleine ville et leurs piles de bétons, les rampes de sortie ou d'accès tels des toboggans gigantesques.
Puis vient le bruit. Le bruit d'une ville est toujours un bruit de fond. Il a envahi vos oreilles en venant de partout à la fois. Ce n'est qu'après un moment que vous y prenez garde et que vous commencez à distinguer les bruits singuliers qui le composent. La circulation, les démarrages au feu vert, les freins à pistons des énormes camions, les sirènes des voitures de police, des pompiers, des ambulances, et des camionnettes de livraison lorsqu'ils reculent, les musiques qui s'échappent des fenêtres ouvertes et des portes des magasins, les exclamations des gens surprises au vol, les conversations lointaines qui sont de purs concentrés d'accent sans parole. Le grondement sourd, presque continue, que vous confondiez avec la circulation et qui provient en fait des avions de lignes, en effet pas si haut que ça."

Thomas Reverdy, *Il était une ville.* Flammarion, 2015, France.

1

Introduction

1

A general background

The objective of this thesis is to measure, understand and explain the interrelations observed in and around the metropolitan area of Brussels by analysing networks of movements. Section 1.1 starts with a description of the studied area before introducing transport geography and spatial interaction models in Section 1.2. Transportation theories are concerned about the movements observed between pair of locations and the characteristics of places from where flows start or end. However, these spatial interaction models are mainly applied on OD matrices provided by censuses, and hence oriented on a long-term evaluation. With the increase importance of big-data (Section 1.3), accurate information of local events appearing in a short period of time are now available, leading to improve the accuracy of our models. Studying movements observed between pair of places, and obtained from large datasets requires new methods. Section 1.4 explains that community detection methods are used to analyse OD matrices obtained from big-data, with no consideration of the characteristics of the places. Mostly, these methods classify all together places that are highly connected, leading to the principle of interaction fields. By comparing these interaction fields to traditional urban models (Section 1.5), this doctoral research aims to bring new insights on transport geography theories. Finally, the general question, as well as the outline of this thesis, are presented in Section 1.4.

1.1 Studied area: Belgium and Brussels

Belgium is a small and densely inhabited European country (11 M inhabitants and 30,528 km^2), with a well-defined urban hierarchy (Vanderstraeten and Van Hecke, 2019) dominated by Brussels. Belgium is made up of three administrative regions separated by a linguistic border: the Flemish (Dutch-speaking), the Walloon (French-speaking, but including nine German-speaking municipalities that constitute the German-speaking Community of Belgium) and the BCR (bilingual). In Belgium, a lot of mainstream social and economic geography is anchored in the context of the so-called city regions ("stadsgewesten"/"aires fonctionnelles urbaines") introduced by Van der Haegen and Pattyn, 1980). These authors observed that initially some basic activities take place outside the traditional city, and hence people operate in a much wider spatial urban environment. As a consequence, it was proposed to deny the use of administrative borders when studying a city and identified the real "cities in new form". Their approach has led to a delineation of functional urban regions around the main Belgian cities, internally organised from the centre to the periphery into the city centre, the central city, the agglomeration, the urban fringe, and the commuter area connected to the city centre (Figure 1.1). Key variables used for delineation were characteristics of places such as population and job density, population evolution, dwelling morphology and percentage of commuters going to the CBD. This methodology was repeated every 10 years, after each new housing and population census (Van der Haegen and Pattyn, 1980; Leemans *et al.*, 1989; Luyten and Van Hecke, 2007). The success of the concept of "city regions" lies not only in the repeated application of a clear methodology since its creation, but also in the fact that it was used and became a standard concept in many academic and administrative geographical analyses in Belgium.

The BCR contains 19 municipalities corresponding to the core of the city (Figure 1.2), and sprawls over a large territory, leading to extensive debate about the exact delineation of the metropolitan area of Brussels (Thomas *et al.*, 2012). The Region in itself is often studied by Belgian geographers (see for instance works made by Dessouroux *et al.*, 2009; Deboosere *et al.*, 2009; Kesteloot *et al.*, 1995; Clerval and Van Criekingen, 2014; Van Criekingen, 2006; Vandermotten *et al.*, 2009b; Van der Haegen *et al.*, 2000, etc.) but also by administrations (amongst other studies made by IBSA - Institut Bruxellois de Statistique et d'Analyse) and the Region itself (see for instance analyses made by de Salle between 2007 and 2009). However, most of the analyses are (1) based on geographical information about places: their characteristics linked with demography, economy, health, etc., and do not often rely on OD matrices (relational datasets), and (2) spatially limited on the BCR.

Determining the delineation of Brussels and its internal composition is not trivial, one influencing the other. This question was raised several times, leading to different spatial extensions due to the variability of data used (Vanderstraeten and Van Hecke, 2019; Hermia, 2005; Dujardin *et al.*, 2007).

At the level of the urban regions, from the city centre to the periphery, the BCR is made up of distinct areas (Figure 1.1 proposes a synthesis of this organisation by mapping some specific structures at the scale of Belgium). First, the city centre corresponds to the urban nuclei, which includes decision and activity centres with the highest concentration of services. Second, the central city corresponds to the area covered by the urban nuclei as well as the highly concentrated neighbourhoods located around this nucleus. Third, the morphological agglomeration is the area covered by the central city and the first crown. Limits of this agglomeration are mainly determined by the contiguity of the built-up (maximal distance of 250 metres between two buildings). Fourth, the operational agglomeration is a simple adjustment of the previous agglomeration, in consideration of the administrative level. Municipalities where at least 50% of the population lives within the morphological agglomeration, belong to this fourth delineation. Fifth, the suburb is found around the agglomeration and its determination is based on several criteria. Sixth, the area covered by the suburbs and the operational agglomeration is called the Urban Region. Seventh, the residential area of alternating migrants includes distant municipalities that are dependent on the jobs offered by the agglomeration. The delineation of this last area is determined by an a priori fixed threshold (proportion of professionally active inhabitants residing in the municipality and having a job in the Urban Region). Finally, the addition of the areas covered by the urban and the residential area of commuters leads to the definition of the urban residential complex. Readers should be aware that although this delineation is largely accepted as a benchmark, it does not correspond to an administrative standard, leading that other extensions based on different criteria and methods are available (Thomas *et al.*, 2017; Verhetsel *et al.*, 2018).

At a local scale, the BCR is divided into different areas (Figure 1.3). The city is articulated around the historical city centre (approximated by the municipality of the City of Brussels), into two main crowns (Deboosere *et al.*, 2009). The first crown was developed in the 19^{th} century and is made up of two subsets: the industrial suburbs (mostly located on municipalities of Molenbeek-Saint-Jean, Anderlecht and Forest) and the middle-class suburbs (mostly located on municipalities of Etterbeek, Ixelles, Saint-Gilles, Saint-Josse-ten-Noode and Schaerbeek). The second more recent crown is also divided into two subsets: the North-West suburb (mostly located on

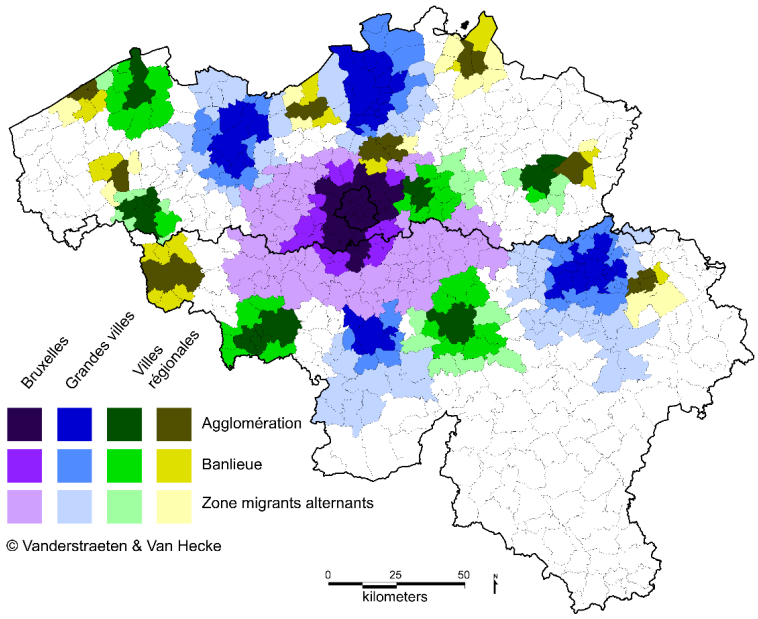


Figure 1.1. The Belgian urban regions as developed by Vanderstraeten and Van Hecke (2019).

municipalities of Berchem-St-Agathe, Evere, Ganshoren, Koekelberg and Jette) and the South-East suburbs (mostly located on municipalities of Auderghem, Uccle, Watermael-Boitsfort, Woluwé-St-Lambert and Woluwé-St-Pierre). As a legacy from the past, this spatial organisation is still strong and stable within the BCR. Differentiation between the various areas mainly depends on the socio-economic status of the residents (Van Criekingen, 2006). The disadvantaged populations are mainly found within and close to the city centre, as well as within places belonging to the first industrial crown developed in the 19th century, while middle classes are mainly residing at the periphery of the city within the second South crown (Dujardin *et al.*, 2008; Van Criekingen, 2006; De Maesschalck *et al.*, 2015).

Analysing the case of Brussels shows that cities are in movements, leading to structures that evolve in space and time. Barthelemy (2016) explains that mobility is probably one of the most important mechanisms that govern the structure and dynamics of cities. Studying these movements is an important facet that may be implemented when considering evolution within space. Transport geography proposes paths for this kind of analysis. This field of research is concerned with mobility, and more particularly how this mobility

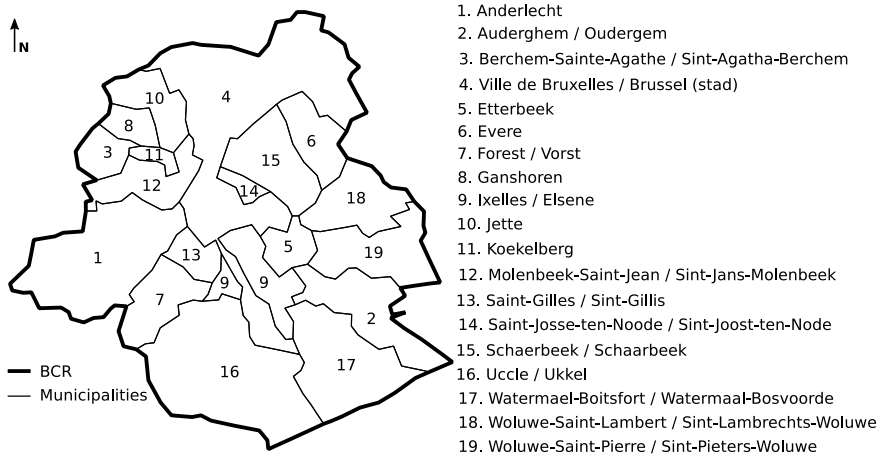


Figure 1.2. Municipalities in the BCR.

is taking place in the context of a wide variety of conditions (Rodrigue *et al.*, 2016).

1.2 Movements, interactions in space

1.2.1 Transport geography theory

"*Transport geography is a sub-discipline of geography concerned about movements of people, freight and information. It seeks to link spatial constraints and attributes with the origin, the destination, the extent, the nature and the purpose of movements*" (Rodrigue *et al.*, 2016).

In other words, transport geography focusses on the locations and the flows observed between pairs of places. On the one hand, it is important to clearly measure the inherent characteristics of each location. Socio-economic conditions, concentration of economic activities, high population density confers, amongst other, a potential supply and/or a demand for resources, products, services or labour. A specific location determines the nature, the origin, the destination, the distance and even the possibility of a movement to be realised. On the other hand, an accurate description of the nature, the purpose and the extent of flows are required. The nature of the goods transported may be a strong impediment to movements. The purpose or the motivation linked with a given movement is an important facet in transport geography: economic, social, transport and communication are only a few

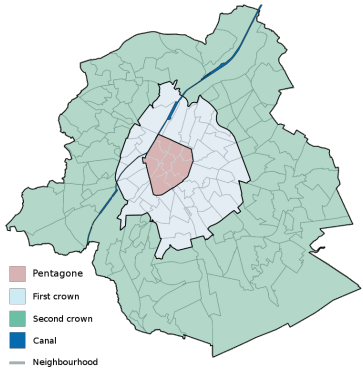


Figure 1.3. A general organisation of the BCR.

examples. The extent of a move has a cost that varies depending on factors, such as the distance and the nature of what is being transported. It is important to notice that the purpose of transportation is to overcome space, which is shaped by a variety of human and physical constraints such as time, administrative divisions and topography. Jointly, they confer a friction to any movement, commonly known as the friction of distance (Rodrigue *et al.*, 2016).

Transport geography may be seen as a network of nodes linked by flows. This network can be considered from a modal perspective, where the edges are an abstraction of routes (roads, rail links, maritime routes) and nodes an abstraction of terminals (ports, rail yards, shopping centres). These networks have an orientation and an extent that approximate their geographical coverage or their market areas. The number of nodes and edges are relevant to express the structure of these transportation networks. Once the various characteristics of the flows and of the interconnected places are theoretically understood, spatial interaction models are useful to analyse and synthesise movements appearing in space.

1.2.2 Spatial interaction models

As developed in the previous Section, transport geography is concerned about movements appearing between places and leading to interactions. Fotheringham (2001) expresses that spatial interaction is a general term for any movement of people, goods or information over space that results from a decision-making process. Modelling spatial interactions generally starts with a matrix of flows between pairs of origins (i) and destinations (j) (Fotheringham, 2001). The principle is to evaluate the number of flows F_{ij}

between any i and j as a function of emissivity of i (measuring the capacity of a place to generate a flow), attractiveness of j (measuring the capacity of a place to attract a flow), and a measure of the distance d_{ij} between the origin i and the destination j (distance measured in kilometres, cost or time). Distance d denotes the difficulty of overcoming spatial separation (O'Kelly, 2015), and mostly expresses the reduction of the probability of spatial interaction with the increase of the distance (Haynes and Fotheringham, 1984; Fotheringham and O'Kelly, 1989; Bailey and Gatrell, 1995). This theoretical framework leads to four different *families of spatial interaction models* (Fotheringham and O'Kelly, 1989; Fotheringham, 2001; O'Kelly, 2015; Haynes and Fotheringham, 1984). First, the *unconstrained model* evaluates interaction when a measure of the distance between i and j is available but when no information is provided for the amount of flow starting at each origin i or ending at each destination j. Based on specific added variables that characterise i (such as population, household income, etc.) and j (amongst other number of jobs, crime rates, the number of parking spaces, etc.), this *unconstrained model* estimates flows between i and j. Second, the *production-constrained model* evaluates spatial interaction when d_{ij} and the number of flows starting on each origin i are known. This kind of model is frequently used in retailing when information about the emissivity of the origins can be easily obtained from census data. Based on characteristics of the destinations, the *production-constrained model* aims to predict the proportion of consumers attending stores located on j. Third, the *attraction-constrained model* is used when d_{ij} and the attractiveness of the destinations j are known. This model is used to understand what are the places that attract most of the flows and hence evaluates what are the characteristics that affect users to choice specific j location? It could also be used to determine the expected numbers of flows reaching each destination j. Finally, the *production-attraction constrained model* (or *doubly constrained model*) is applied when the number of flow starting at each origin i or ending at each destination j and d_{ij} are supposed known. Transportation applications mostly use this model to predict the distribution of trips between i and j.

It is important to notice that distance is the most significant impedance factor when evaluating interactions between places. However, distance is not homogeneous in space and discontinuities may appear, corresponding to borders that affect interactions (Nijkamp *et al.*, 1990; Dujardin, 2001). Absolute (such as a river) or relative (the topography that influence the path) are hence barriers that have to be considered within spatial interaction models. As an example, it is largely accepted that two places belonging to the same territorial entity (the same linguistic region for a Belgian case study)

exchange more between them than with places allocated into different territorial entities, (see Grasland, 1999). This effect could result from a high cost (money, time or psychological) linked with the crossing of the barrier (Nijkamp *et al.*, 1990; Grasland, 1999) or with a bad knowledge of the relations existing in another territorial entity, as regard the emissivity and attractiveness induced by this spatial gap (Grasland, 1999).

Spatial interaction models are mostly applied to OD matrices provided by censuses, based on theoretical concepts, assumptions or proxy to evaluate the characteristics of the places, and are mainly oriented to a long-term planning. Currently, with the increasing importance of big-data within the daily life of people, a large amount of information is now available to describe the choices of the population located in particular places, as well as a short-term estimation of the dynamics observed between pair of locations. Section 1.3 describes, defines and evaluates these Big-Datasets.

1.3 Big-data

Since 1980, researchers deal with the concept of ICT (Information and communication technology) that corresponds to any communication (telephone lines, wireless signals, etc.), computer systems, and to the necessary software that enable users to access, store, transmit, and manipulate information. ICT allows people (or objects) to be connected all together and hence it modifies the way they communicate and interact. More recently, mobile phone devices connected to the internet (smartphone) strengthen interactions made by people. More than just facilitating communications, these devices are equipped with several tools (GPS, cameras, brightness sensor, etc) that help users to take "advantage" of their environment and to enhance the way they interact with their social network.

Thereby, ICT produces a large amount of newly varied data (location, social network, sensors, pictures, etc.) that are hence collected, stored and analysed: these massive kinds of datasets are called big-data (Graham and Shelton, 2013). Several definitions were developed, leading to the "V theory" which is used to describe each of the characteristics. Amongst other, big-data are: (1) big in Volume - Terabytes, (2) big in Velocity - real time, and (3) big in Variety (Boyd and Crawford, 2012). These new data sources are often messy, consisting of data that are unstructured, collected with no quality control, and frequently accompanied by no documentation or metadata (Miller and Goodchild, 2015). In addition, these datasets are mostly collected in real-time, meaning that unplanned events can be captured. Hence, questions have to be made, linked with the Veracity (are we dealing with information or disinformation?), the representativeness (what are the sample

processes - are the data exhaustive?) or the Value (can we value results?). Although these several issues, big-datasets are a major opportunity, and especially in the era of urban science and planning. Batty (2012) explains that initially urban sciences were mainly centred on "radical and massive changes to cities over the long-term, with little concern for small spaces and local movements". This large amount of data hence allows to analyse local events appearing on shorter terms.

Dealing with big-data is challenging, not due to the Volume of the dataset but due to the high Velocity and Variety (Longley *et al.*, 2015). Hence, there is a need to transform these heterogeneous data collected from diverse data sources into a clear and structured information useful to obtain results (Cuzzocrea *et al.*, 2011). This quick way to obtain a large variety of data questions the "ability and usefulness of our theories and explanations when we can measure and describe so much, so quickly" (Miller and Goodchild, 2015). These authors refer to two major theories: *theory-driven* and *data-driven*. On the one hand, *theory-driven* is based on the creation of a conceptual theoretical framework, based on assumptions and choices. Data are implemented and used to confirm if the model is useful. On the other hand, *data-driven* models use the data as a cornerstone for developing theories and models.

Whatever the theoretical approach chosen, new techniques are required to evaluate the large OD matrices computed from these datasets. Amongst others, community detection methods succeed in grasping interactions appearing between places, by grouping together locations that are tightly connected. General explanations of these methods are presented in Section 1.4.

1.4 Community detection methods

A graph or a network is a set of nodes that are linked together by (un)weighted edges. A node could represent a person, a place, a phone antenna or another specific item. A given weight is associated with each edge, corresponding to the intensity or the direction of the interaction, like the number of people moving between two municipalities or the number of calls made between two antennas. Some specific nodes within a given network are more densely connected between them than with the rest of the network, leading to the definition of "communities of nodes". Without any consideration of the characteristics of the nodes, these community detection methods reveal the internal structure of the network and help to visualise the hierarchy of the relations appearing between entities (Fortunato and

Lancichinetti, 2009). Different families of methods were developed to highlight these strongly connected nodes: amongst others, Traditional, Divisive algorithms and Modularity-based methods (Fortunato and Lancichinetti, 2009).

Traditional methods refer to (1) *graph partitioning* that requires an a priori number of communities, and (2) *hierarchical clustering* based on bottom-up (Agglomerative algorithms) or top-down (Divise algorithms) approaches and that may be illustrated by means of dendrograms and could sometimes require specific stopping criterion (Fortunato and Lancichinetti, 2009).

The philosophy of the Divisive algorithms is to find the edges that connect nodes of different communities, and to remove them, leading to the creation of disconnected groups (Fortunato and Lancichinetti, 2009). The algorithm proposed by Girvan and Newman (2002) is considered as a cornerstone of a new era of research in the field of community detection (Fortunato and Lancichinetti, 2009). It is based on several important concepts - edges centrality or betweenness - that respectively estimate the importance of edges within a network and the frequency of the participation of edges to a process (Newman and Girvan, 2004).

Finally, Modularity-base methods correspond to several algorithms based on the Modularity optimisation (or a modified version of the modularity). This modularity was firstly defined as a stopping criterion for the algorithm developed by Girvan and Newman (2002) and then became an essential element of many clustering methods to quantify the quality of the partitioning computed (Fortunato and Lancichinetti, 2009).

Compared to spatial interaction models (Section 1.2.2), these methods are not concerned about the characteristics of the places (nodes of the network). However, characterising places of the origin and the destination of flows stay a major question in spatial interaction models. Urban models presented in Section 1.5 propose paths for a better description and understanding of the organisation of urban areas.

1.5 Urban models

Urban areas are, by definition, complex systems where multiple realities are embedded: demographics, economy, environment and transport are different layers that are closely interrelated (Batty, 2013; Barthelemy, 2016). It leads to many positive externalities attracting people and activities. Understanding urban processes is hence a major challenge to better manage and improve future daily urban lives. Doing so requires to accurately define what is a city and what is its spatial extension (see Section 1.1). Not having this theoretical framework would lead to major issues that under or over-estimate the

delineation of urban areas and hence lead to miscalculations or wrong statistics (Coombes, 2013; Thomas *et al.*, 2018).

Modelling the urban environment based on characteristics and hence understanding cities is not a novel challenge (Delloye, 2018, for a literature review). Since decades, researchers model what processes are behind urban realities (Hall *et al.*, 2012). Works made by urban sociologists at the University of Chicago were particularly influent within geography, and more particularly the models developed by Burgess in 1925 (concentric zonal model) and Hoyt in 1939 (sector model).

Burgess modelled the industrial city as made of several concentric areas organised around the Central Business District (CBD), which corresponds to the city centre. This CBD is dominated by the commercial function surrounded by a transitional zone where industry is mixed with poor housings. Clearly, this model identifies a rise in the functional segregation, with an increasing separation of home and work. Furthermore, separation between the socio-economic situation of people is also modelled: social status and wealth increase from the city centre towards the urban features (Burgess, 1925).

At the opposite, Hoyt's model argues that residential areas could be better understood as a pattern of sectors instead of a concentric organisation. Hoyt developed the idea of sectors of high-rent areas and he suggested that high status sectors are found along transportation network radiating out from the centre and away from industrial areas. In addition, wealthier populations have a higher propensity of moving out of the centre (Hoyt, 1939).

Two supplementary models were developed in the same vein. Firstly, two geographers, Harris and Ullman describe the city layout in 1945. They made the argument that the CBD becomes less important than the rest of the city: it only corresponds to a local point within the urban area. This theoretical development was made during the increasing interest of people to own their private automobile. The distance cost hence became smaller and residents could travel for long distances, making suburbs more accessible from cities. At the opposite of the two previous models, Harris and Edward explained that cities have to be seen as a multiple-nuclei model instead of a simple core periphery model. They made the distinction between nine different urban sections with their own function (Central business district, Light manufacturing, Low-class residential, Middle-class residential, Upper-class residential, Heavy manufacturing, Outlying business district, Residential suburb and finally, Industrial suburb).

The second supplementary model developed by Vance improves the multiple nuclei model in 1964 by proposing the urban-realms model. Based on the San Francisco's urban ecology, he summarises the economic processes, and his model suggests the existence of various "realms". These are "self-sufficient

urban areas with independent centres" and have to be examined based on five criteria (physical geography, size of the metropolitan area, strength of the economic activity within each of the realms, the internal accessibility of each realm and the inter-accessibility across the individual suburban realms). This model succeeds quite well to explain the urban growth as well as the de-localisation of some specific activity from the CBD to suburbs.

It is important to specify that, on the one hand, these urban models aim to drastically simplify the reality, and on the other hand, to capture enough of the social reality for comprehension and decision-making. It leads to simple models that were not efficient enough to understand all the urban facets, leading to the complex theory (Batty and Torrens, 2001). This complexity refers to each dynamic which enables their elements to transform in ways that are unpredictable. It leads to an impossibility to create models that efficiently represent the reality because of the boundless complexity (Batty and Torrens, 2001). Pumain (2003) expresses that geographical objects (cities, regions, networks of cities) should no longer be explained by a simple "biographical narrative" but as one of the many possibilities of a set of complex interactive processes.

1.6 Objectives of this contribution

This thesis aims to bring a new look on flows measured between pairs of origin i and destination j. These flows may be seen as interaction appearing between places and hence corresponding to movements. As introduced in Section 1.2, movements are studied by transport geographers or other spatial interaction models concerned about the characteristics and the interaction of places. These methods were mainly applied on traditional OD matrices provided by Statistics Services (Fujita and Thisse, 2013; Glaeser, 2011), and results were hence limited by the quality of the data (restricted assumptions, privacy issues leading to aggregation problem, etc.). However, new large datasets (big-data) are currently offering a huge amount of information at very fine scales (mostly individual) and in real-time (Section 1.3). Analysing this pile of data requires new computational techniques to obtain quick robust results. When dealing with OD matrices provided by Big-Dataset, community detection methods (Section 1.4) are useful to find how places are connected and what are their extents in space. With this advance in graph theory, one insists on the need for more deeply study the relation embedded within cities (Rozenblat and Melancon, 2013). Intra-urban analyses of the economic, social and political networks should hence generate clear and useful information on the processes and strategies reinforcing the inter and intra-urban connections. This would improve our knowledge of the closeness of the multiple agents

(economic and population) connected within urban areas, leading to the principle of interaction fields.

The general objective of this doctoral research is hence to measure, understand and explain the interrelations observed in and around the metropolitan area of Brussels by analysing networks of movements constructed from large (un)conventional OD matrices. To this end, and without any consideration of the characteristics of places, this thesis evaluates if community detection methods (Section 1.4) may provide useful spatial partition of the territory within movements of people, freight and information. As already explained in Section 1.2.1, transport geography seeks to link spatial constraints and attributes with the origin, the destination, the nature, the purpose and the extent of movements (Rodrigue *et al.*, 2016). By comparing the interaction fields detected in Big-(relational)-Datasets (Section 1.4) to urban models (Section 1.5), we propose a new view to link interactions to the characteristics of the nodes, leading to a new approach to deal with transport geography and hence approximating the *attraction-constrained model* (Section 1.2.2).

Although theoretical frameworks are well understood within the scientific literature, analysing the structure and the organisation of networks at different levels of resolution are unfortunately not enough handled. This thesis proposes to overpass this lack, by implementing a multi-scale component within the Louvain Method as well as proposing graphical and cartographic outputs. It leads to explain the geographical extent of the computed communities: how they are organised in space, how strong they are embedded within the global hierarchy of the network and how these groups overlap. Hence, more than just detecting the spatial footprints and the composition of these various groups, the limits are in depth discussed in a spatial planning point of view with a major transversal question: do the borders of communities correspond to the administrative ones? In other words, are our administrative units able to integrate the realities appearing in space?

To accurately describe and understand how interactions are organised within the Metropolitan area of Brussels, this doctoral dissertation is divided into six parts, where Parts I and II introduce theoretical concepts and methodological framework. Part III, IV and V respectively discuss movements of people, freight and information. Finally, the last Part VI concludes this thesis. Figure 1.4 proposes a schematic outline of the thesis. The three types of movements studied are developed with their significations and their purposes.

Part I introduces the thesis and the general empirical, methodological and

geographical contexts. The motivation, the objectives and its structure are developed. In particular, Chapter 1 firstly proposes some explanation about the spatial organisation of Belgium and Brussels, as well of some perspectives. A conceptual presentation of transport geography and spatial interaction models is secondly made to better introduce urban and interaction models.

Part II is composed of two chapters and presents the methodological framework of the community detection method. More than just applying "black-box" tools, this contribution aims to present the methods used. By pursuing the development of methodologies able to analyse networks from large databases, while evaluating their biases, their uncertainties and their spatio-temporal evolution, the quality of results will be improved, leading to a better understanding of Brussels. A high care is hence done on the methodological aspects, and more accurately on the sensitivity of the results to (1) the methods and the parameters chosen, (2) the hierarchy of the spatial partitions obtained and (3) the robustness of the results. Consequently, Chapter 2 mainly focusses on testing the sensitivity of the Louvain Method (one of the most applied and famous method to detect communities in large networks) to these various methodological and geographical aspects. In Chapter 3, the Louvain Method is confronted to a hierarchical clustering method (Section 1.4) that classifies places based on their interrelations. Analyses in Chapter 2 and 3 are applied on two "traditional" datasets collected from the last Belgium Census (2011): the commuting and the migration (changes of residence) movements. Part 2 is hence drawing the first basins resulted from the movements made by the Belgian population. These spatial structures will be analysed deeper within the third Part.

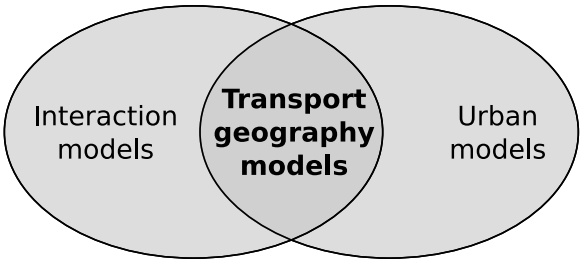
Part III presents spatial partitions detected in the movements made by people. It aims to refresh the geography of Belgium (as a benchmark) and the Metropolitan area of Brussels, based on datasets that are different in nature. On the one hand, Chapters 4 deals with commuting and migration movements provided by the last Belgian census, leading to delineate and compose job and residential areas based on theoretical movements (see Figure 1.4). On the other hand, the Chapter 5 focusses on the analysis of an "unconventional" Big-Dataset: train schedules requests made on internet and that represent potential movements. This Chapter also raises few questions regarding potentialities of data scrabbled on the internet: can they be used as a proxy of transport demand?

In Part IV, Chapter 6 explores the real movements of freight in Belgium, approximated by the circulation of trucks. Based on GPS data points emitted by GPS trackers during the circulation of trucks, the objective is to delineate

the hinterland of Brussels by revealing only one of the dimensions of the economic polarisation of the city. The major contribution of this analyse resides in the representativeness instead of using samples or surveys. This Chapter deals with a Big-Dataset that contains every position of each truck circulating in Belgium. By computing the OD of each truck trip, the places visited by trucks are determined in real time. Analyses are made at two different scales: firstly, Belgium is considered as a benchmark helping us to secondly approximate the place of Brussels within Belgium and what are the close or distant places that are polarised by Brussels.

Part V deals with three analyses based on mobile phone calls, and presented in Chapter 7. It evaluates real movements of information and hence approximates social interactions between people. The main idea behind this Chapter is to approximate and measure the three main aspects of the urban complexity: interactions, activities and characteristics. Firstly, by means of more than 13 million mobile phone calls and the Louvain Method, antennas that are closely connected are classified all together, revealing major communication patterns within the Metropolitan area of Brussels, their spatial footprints and their hourly evolution in time. Secondly, the way people are active within space is approximated by the variation in the time of the activity of antennas. The ones that have the same hourly variation of calls are classified all together, leading to delineate residential, economic or mixed areas. Thirdly, conventional socio-economic variables selected in the last Belgian Census are used in a Ward classification method to group all together places having similar characteristics. Finally, this Chapter concludes with a general and concluding partition of the Metropolitan area based on the three previous analyses, leading to integrate the characteristics of places within movement of information through space. A discussion about the usefulness of mobile phone calls to describe the urban structure is also presented.

Finally, Part VI comes back on the general objective of this contribution: measuring, understanding and explaining interactions appearing in and around the metropolitan area of Brussels, leading to new views on transport geography. Chapter 8 synthesises the most striking results into clear messages integrating the diverse components of the urban complexity and transport geography. More than just proposing a synthesis, a critical assessment is made, as well as a development of the several paths that may be followed for further research. In addition, the implications of this contribution are highlighted, not only for academics but for public and private sector stakeholders as well.



Type of moves	Dataset	Signification	Purpose	Chapter
PEOPLE	Commuting	Theory	Economic	2 - 4
	Migration	Theory	Economic Social	3 - 4
	Train schedules requests	Potential	Transport	5
FREIGHT	Trucks circulation	Real	Transport Economic	6
INFOR-MATION	Phone calls	Real	Commu-nication	7

Figure 1.4. Schematic outline of the thesis.

II

Methodological framework

2

The multi-scale Louvain Method to detect communities: robustness tests

New methods are nowadays developed for analysing the enormous masses of data provided by the many ICT devices. Among these, the detection of communities in large networks has become common practice in socio-spatial analyses.However, a much greater care should be taken regarding specific methodological issues such as the choice of the value of some parameters or the intrinsic characteristics of the spatial data. In this view, this Chapter tests the sensitivity of the results of the multi-scale Louvain Method to the values of the resolution parameter (Section 2.3.1) and to two crucial spatial data issues: the inclusion of internal loops and the delineation of the study area (see Section 2.3.2 and 2.3.3). We also compare the communities obtained by the Louvain Method with those obtained by the Infomap method.

Sensitivity analyses are conducted for commuting movements in and around Brussels. Results reveal that job basins are slightly affected by the method or by the spatial issues, while the resolution parameter plays a major role in the results and their interpretation in terms of urban geography.

This Chapter is based on works already published in the *Journal of Geographical Systems* (see Adam *et al.*, 2018b)

2.1 Introduction

Urban areas concentrate a large number and a great diversity of people, buildings, activities and functions. Consequently, interactions are numerous, intense, complex and vary in nature (Scott and Storper, 2015; Fujita and Thisse, 2013; Anas *et al.*, 1998). Quantitative geographers often measure urban realities as a set of characteristics X associated to places i and/or as a set of interactions W between places i and j. Both data types have led to the development of methods and theories for determining the delineation and the composition of urban areas. While urban factorial ecology was early and rapidly accepted as a standard approach for clustering places with similar characteristics X_i that fitted well with classical urban models (Hoyt, Burgess, see e.g. Kitchin and Tate, 1999), there is less methodological consensus about the delineation of the *interaction fields* W_{ij} (see Fujita and Thisse, 2013) for defining urban boundaries or for identifying specific intra-urban basins such as job basins (Duranton, 2015; Dujardin *et al.*, 2007). A job basin pertains to only one facet of the interaction fields. It corresponds to a group of places having a strong economic relation or linked between by intense commuting (residence place of work) and it is generally dominated by one or to multiple centres (Géoconfluences, 2017). Job basins are generally delineated on the basis of origin-destination flow matrices (O'Kelly *et al.*, 2012; Li and Farber, 2016).

The delineation of *interaction fields* has been the subject of numerous publications, covering methods from the simple a priori defined threshold up to interaction models with sophisticated parameters (for a critical review; Haynes and Fotheringham, 1984; Jacobs, 1969; Dujardin *et al.*, 2007). For example, the interactions between places can be classically clustered based on graph theory and hierarchical clustering; many methods, such as the Anabel method developed by INSEE in France (insee, 2015) have been applied in regional research and in public administration. This method iteratively groups municipalities together based on the highest flow found in the transition matrix. At a larger scale, communication patterns between regions have been modelled by analyzing phone calls with techniques such as Intramax and IPFP approaches (Fischer *et al.*, 1993).

Although these methods show interesting results in terms of spatial organization, the partitioning is deterministic and may be sensitive to the thresholds or the choices to be made during the implementation. Instead of producing dendrograms that visualize the *interactions patterns*, we aim to detect the organization of a network into groups of nodes (Fortunato, 2009). Indeed, a network is a set of nodes linked together by edges that can be weighted (for instance by the number of commuters between places), and a

community is a set of highly connected nodes that are more intensively linked together than with the rest of the network (Danon *et al.*, 2005; Fortunato and Lancichinetti, 2009). Such communities at different resolution levels can be identified by means of a new generation of algorithms. Because of their simplicity and the availability of new data sources (sensors, Big-Data, ICT data), the interest in these methods has increased with the application in various research fields such as physics and mathematics (for a general review, please, read Fortunato, 2009), socio-economics (Beguerisse-Díaz *et al.*, 2014; Cerina *et al.*, 2014; Crampton *et al.*, 2013; Onnela *et al.*, 2011; Papadopoulos *et al.*, 2011), and spatial planning (Blondel *et al.*, 2010; Chi *et al.*, 2016; Montis *et al.*, 2013; Ratti *et al.*, 2010; Jones *et al.*, 2016).

This chapter relies on one of the most widely used community detection methods (the multi-scale Louvain Method) and tests its usefulness for delineating interaction fields as well as regional science topic; and more precisely, job basins within the Brussels metropolitan area. The Louvain Method does not require any information about job centres and allows researchers to analyze communities of commuting flows without relying on spatial information. No a priori centre is here defined and spatial information is not used as an input for these community detection methods (such as the localization of places).

Moreover, only the total flow between places is considered here, and not the direction of flows. Let us consider two locations, m and n: if the flow between m and n is equal to 50 and the flow between n and m corresponds to 90, we here consider only the total exchange (140). The contribution of this chapter is threefold: to test the robustness of the Louvain Method to some crucial inputs; to illustrate the potential of the method for urban delineation in the Brussels metropolitan area; and to, finally, compare the results with an alternative method, called Infomap. For this purpose, sensitivity analyses are performed with respect to, the resolution parameter (Reichardt and Bornholdt, 2004), the effect of the inclusion (or not) of internal loops in the data set, and, extending of the study area. Computations are based on commuting flows in and around Brussels (Belgium) and results are further commented in terms of well-known Belgian geographical issues. We show that distance still matters (Ioannides, 2012) and that nearby places are more strongly related than are distant places. This chapter is organized as follows: methodological issues and data choices are dealt with in Section 2.2.1, Section 2.3 presents and discusses the results, while Section 2.4 provides conclusions.

2.2 Methodology and data

2.2.1 Community detection

Many recent community detection algorithms are based on modularity maximization (Newman and Girvan, 2004) and are applied to networks, allowing the detection of structures consisting of nodes linked together by (un)weighted edges. The higher the weight between two nodes, the more they are related and the more the modularity will increase by classifying these two nodes in the same community. A community is hence a group of tightly connected nodes. To quantify the quality of a partition into communities, modularity compares how densely the nodes are connected within the communities in comparison to a random null model. Indeed, modularity is the sum of the weighted edges inside each community minus the number of expected edges between the same nodes in a random null model with the same degree of distribution and the same number of edges (Newman and Girvan, 2004). Modularity therefore quantifies how densely the communities are connected, with respect to the density of edges that would be expected of the same group of nodes in a null model where nodes are linked randomly while maintaining the same degree.

Using modularity as an objective function produces an unpredictable number of communities of homogeneous size where the size is the sum of all the weighted edges classified in the community and does not refer to the number of nodes (Fortunato and Barthélemy, 2007; Delvenne *et al.*, 2010). Delvenne *et al.* (2013) reinterpret modularity (Q) by writing the following simplified Equation 2.1:

$$Q = Diversity - Cut, \tag{2.1}$$

$$Diversity = 1 - \sum_{c \in P} \left(\frac{in[c]}{2m} \right)^2, \tag{2.2}$$

$$Cut = 1 - \frac{in[c]}{2m}, \tag{2.3}$$

where P is a partition (a set of communities c), $in[c]$ is the sum of all the weighted edges totally included in community c, and m is the sum of all the weighted edges of the network. More precisely, determining the organization of the network into communities, modularity (see Eq. 2.1) aims at minimizing the Cut size (the total weighted edges between communities) while maximizing the $Diversity$ index (the number of communities of homogeneous size detected) that are two opposed objectives (see Delvenne *et al.*, 2013). Minimization of the Cut size would lead to one single community including

all the nodes of the network, while the maximization of the *Diversity* index provides a partition of nodes into one community. A main drawback of the modularity function is the *resolution limit* (Fortunato and Barthélemy, 2007). This limit is due to the *trade off* made between maximization of the *Diversity* and minimization of the *Cut*. In some cases, depending on the number of nodes and edges, some small communities are incorporated into larger communities although theses are only linked by one edge.

The Louvain Method is a greedy heuristic that searches for a maximum modularity partition of the nodes in successive steps, which are briefly described as follows (for more details, see Blondel *et al.*, 2008). The method starts with an initial partition where each node is a community: the number of communities (Z) equals the number of nodes (N). Then, the method merges the nodes that are highly related in successive steps and creates communities that increase the value of the modularity. These merges are made until a local maximum of modularity is reached.

This chapter proposes to deal with a modified version of the Louvain Method[1] to evaluate the multi-scale component of the modularity. Reichardt and Bornholdt (2004) introduce a resolution parameter ρ in Eq. 2.4 to influence the maximization of modularity. This modification allows the detection of partitions composed of communities of different sizes. The modified modularity Q_ρ is given by

$$Q_\rho = Diversity - \rho\ Cut. \tag{2.4}$$

For Lambiotte (2010), the parameter ρ is seen as a value of time (cf. random walker) for modularity to group together the highly linked nodes. The effect of the use of different values of ρ is illustrated in Fig. 2.1. With a zero value, there is no gain of modularity and the communities detected are those of the initialization step ($Z = N$). With a high value of ρ, modularity always increases with each merge and the Louvain Method detects one single large community including all nodes ($Z = 1$). When intermediate values of ρ are used, Z ranges between N and one. Let us note that when ρ is equal to one, the partition corresponds to the one detected endogenously by the original Louvain Method.

Although the method can detect communities of different sizes, the problem of the resolution limit is still present and using modularity as a general indicator of the *quality of the partition* (Lambiotte *et al.*, 2014; Steinhaeuser and Chawla, 2010) only gives a value of quality for the entire partition and not for each community. To describe each community independently, we suggest to calculate a measure evaluating the *Cut* (see Eq. 2.4) which we here call *compactness*. The

[1]For the sake of clarity, the term Louvain Method used in this contribution refers to the multi-scale Louvain Method

compactness F of a community c is defined as:

$$F(c) = \frac{in[c]}{all[c]}, \tag{2.5}$$

where $F(c)$ evaluates the ratio between the sum of the weighted edges included within a community c and the total weighted edges of this community c (Fig. 2.2). It ranges from zero to one; where a value close to zero expresses that a community has more edges going to other communities, and a value close to one reveals that the community is self-centred and densely connected. This indicator provides information about the connections of a community to the rest of the network.

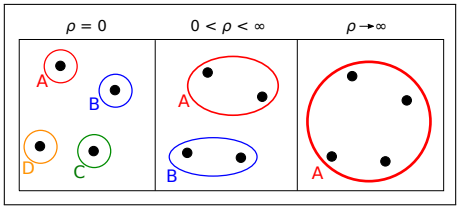


Figure 2.1. Partitions detected using different values of ρ, in multi-scale modularity (A, B, C, D communities).

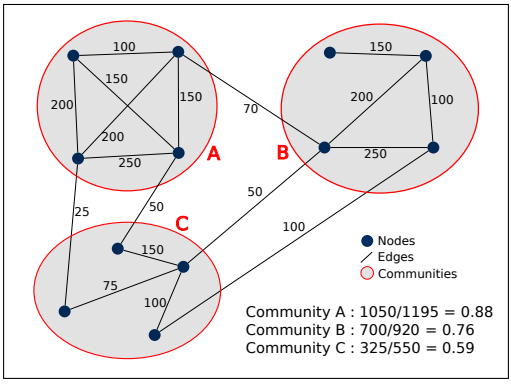


Figure 2.2. Compactness of communities.

Last but not least, let us point out three major characteristics of modularity. First, all edges have an influence on the detection of communities without taking into account the spatial location of the nodes; only the

relational information is used. Second, the method detects endogenously the number of communities and the size of these structures tends to be homogeneous. Finally, the partitions obtained are sensitive to the order in which the nodes are considered by the method. Thus, if the order of the nodes is not changed, the Louvain Method gives the same deterministic results; sensitivity to the change of order is hence a way of testing the robustness of the results. In order to detect robust communities we run the method 1,000 times, randomly shuffle the order of the nodes and extract the dominant structure from the multiple partitions obtained. The applied methodology (see Appendix for more details) allows to calculate the instability of the classification of the nodes. The instability corresponds to the number of times that the nodes are not classified in their dominant community during the 1,000 runs of the Louvain Method. Instability moderates the classification of municipalities into communities. Some municipalities do not have a clear classification, indicating a tendency in terms of commuting movements are fuzzier.

2.2.2 Data

Let us consider commuting movements in 2011 (Census-2011, 2016). Due to data limitation, there is no information on commuting flows from locations outside of Belgium. The nodes represent the 589 municipalities of the country and the edges are weighted by the number of commuters between pairs of nodes. In order to avoid this discussion of the limits of the study area, this chapter focuses on the former Province of Brabant; that is to say, the 19 municipalities of the BCR and 92 municipalities located around the BCR, in Flanders (north of the linguistic border) and in Wallonia (south) (Fig. 2.4).

The BCR offers job opportunities attracting commuters from all over the country (Riguelle *et al.*, 2007). More specifically, many workers living in the province of Brabant commute to the BCR for work: 32% of workers living in Flemish Brabant work in the BCR (more than 0.32 M), 35% of workers living in the Walloon Brabant commute to the BCR (more than 0.13 M) and finally, 87.4% of workers living in the BCR work in the region (more than 0.32 M) (Census-2011, 2016).

By analyzing the communities of commuters in the province of Brabant, we aim to delineate job basins in and around Brussels. As explained in the introduction, a job basin is a group of municipalities linked together by strong commuting (residence - place of work) (Géoconfluences, 2017). Fig. 2.3illustrates for two municipalities (one of the 19 within the BCR, the city of Brussels, and the municipality of Leuven) the proportion of workers who commute from other municipalities to reach these specific places for work. On the one hand, a large proportion of workers in the city of Brussels

live in the numerous municipalities located within or close to the BCR, as well as in the Flemish and the Walloon municipalities. On the other hand, workers in the municipality of Leuven have a residence close to this centre in the eastern part of the former province of Brabant. Fig. 2.3 illustrates two distinct spatial residential choices based on the place of work.

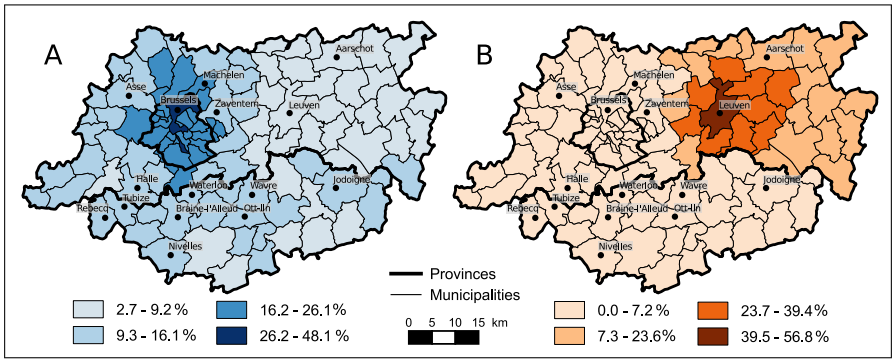


Figure 2.3. Proportion of in-commuters working in the municipality of Brussels (A) and in the municipality of Leuven (B).

As already explained above, the Louvain Method delineates communities without a priori definition of a centre. To quantify the centrality of places, the weighted degree of each node is calculated. The weighted degree of a municipality i is the sum of all the weighted edges of this node (Freeman, 1978). In other words, the weighted degree represents the total number of commuters arriving at and leaving from a municipality as well as the intra-municipal workers. These intra-municipal workers (internal loops) correspond to what is called here "self-loops". As expected, Fig. 2.4 shows that the highest weighted degree values are located within the BCR as well as in the secondary centres (Leuven, Zaventem, Wavre) in accordance with the urban sprawl mechanism observed in that area (Riguelle *et al.*, 2007). To evaluate the importance of the self-loops in this commuting network, we calculate the weighted degree of the municipalities when all the self-loops are erased and we notice that the spatial distribution of the weighted degree is similar to the distribution obtained with the inclusion of self-loops. Table 2.1 shows that the municipalities located in or close to the BCR (Bruxelles and Zaventem) have a smaller loss of weighted degree (23.0% and 17.8%) than the municipalities further away such as Leuven (46.6%). Indeed, municipalities located close to or within the BCR are less influenced by intra-municipal workers because many workers commute from places outside of the BCR (Cheshire, 2010; Verhetsel *et al.*, 2010; Thisse and Thomas, 2010b).

Table 2.1. Weighted degree size-ranked municipalities within the former province of Brabant.

Municipality	With self-loops		Without self-loops	
Bruxelles	189488	(1)	145910	(1)
Leuven	91837	(2)	49931	(2)
Anderlecht	69127	(3)	49267	(3)
Schaerbeek	65718	(4)	48310	(4)
Ixelles	59274	(5)	43166	(5)
Uccle	50859	(6)	32569	(6)
Molenbeek Saint Jean	43228	(7)	32320	(7)
Zaventem	38405	(8)	31557	(8)
Woluwe Saint Lambert	37867	(9)	28625	(9)
Saint Gilles	35559	(10)	26899	(11)
Etterbeek	34517	(11)	27229	(10)

To categorize the municipalities in two arbitrary groups, the Jenks classification is applied on the weighted degree distribution. The classes are defined according to the distribution of the data values and the limits of the classes are based on break points of the distribution; it minimizes the internal variance of classes and maximizes the variance between classes (Longley *et al.*, 2010). The municipalities characterized by the highest values of weighted degree are represented on the figures with black dots located at the centre of these places. In Fig. 2.4, the highest values of weighted degree (Jenks classification) are highlighted by their names. They correspond to employment centres and edge cities (Riguelle *et al.*, 2007) but not all small urban centres appear (Tienen, Nivelles, Asse, etc.).

2.2.3 Sensitivity analyses

First, we try to understand and measure how far ρ affects the partitions. Second, we evaluate the impact of the inclusion (or not) of the self-loops (intra-municipal commuters) in the data on the partitions and, third, we estimate the effect of isolating the province of Brabant from the rest of Belgium (that is to say, considering 111 municipalities instead of 589). Each analysis is run 1,000 times and the order of nodes is randomly shuffled for each iteration. These two points guarantee the robustness of the structure and allow the detection of the unstable municipalities (places that are not always classified in the same community through the 1,000 runs). Each municipality is hence characterized by a percentage, representing the relative number of times it is classified in its dominant community. A value of 100% corresponds to a fully

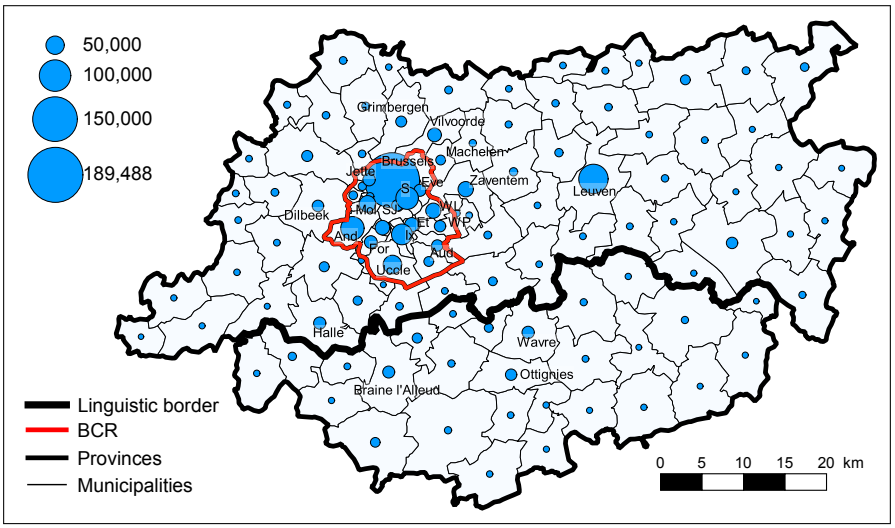


Figure 2.4. Weighted degree of the municipalities within the former province of Brabant (including self-loops).

stable structure. We here decide that the municipalities classified less than 95% with their dominant community are represented on the maps with white hatches to allow comparison over the various maps.

To quantify the comparison between two partitions, we use the Normalized Mutual Information criterion (*NMI,* Strehl and Ghosh, 2002), defined as

$$NMI(\omega, \upsilon) = \frac{I(\omega, \upsilon)}{(H(\omega) + H(\upsilon))/2}, \tag{2.6}$$

where ω and υ are two partitions, $I(\omega, \upsilon) = H(\omega) + H(\upsilon) - H(\omega, \upsilon)$, with $H(\omega)$ and $H(\upsilon)$ representing the Shannon entropy of the partitions ω and υ, respectively, while $H(\omega, \upsilon)$ is the Shannon entropy of the joint probability. NMI returns a value of similarity that ranges between zero and one. A value close to one expresses a very good similarity between the two partitions.

2.3 Results

The sensitivity of the Louvain Method is first presented for the variation of ρ, secondly for the inclusion or not of self-loops, and finally for using two different study areas. For the sake of clarity, we only compare partitions with a same number of communities in Section 2.3.2 and 2.3.3,.

2.3.1 Effect of parameter ρ

We apply the linearized stability procedure developed by Delmotte *et al.* (2011) to select a set of values of ρ used for the sensitivity analysis. This method tests the robustness of the communities found for various values of ρ (range from 0.1 to 10.0 with a step of 0.1) by applying the Louvain Method several times. The stability method thus provides multiple partitions for each ρ. Delvenne *et al.* (2013) explain that a robust community structure is persistent through the multiple partitions found for each ρ and can be observed (1) when the *variation of information* (VI) is minimal between the partitions made with a specific value of ρ or, (2) when the curve of the number of communities detected has a plateau shape. VI between two distinct partitions ω and υ is defined as:

$$VI(\omega, \upsilon) = H(\omega, \upsilon) - I(\omega, \upsilon). \tag{2.7}$$

Hence, a low value of VI indicates that two partitions are very close to each other and the structure of the communities to be robust to varying ρ. Let us note that NMI and VI (Eqs. 2.6 and 2.7) both aim to test the mutual information shared by two partitions (minimal for the VI and close to one for the NMI).

Figure 2.5 illustrates the use of the stability method for the former province of Brabant including the intra-municipal movements (self-loops). In this case, we propose to select values of ρ that provide different characteristics: two values with the smallest VI (0.7 and 1.5), two values on a constant slope (0.5 and 0.9), and two breaks of the slope (1.0 and 2.0). For each of these values, communities are detected, analyzed and mapped (Fig. 2.6). The color used to discriminate the communities are chosen arbitrarily and the municipalities distinguished by white hatches correspond to places that are classified as less than 95% of their dominant community.

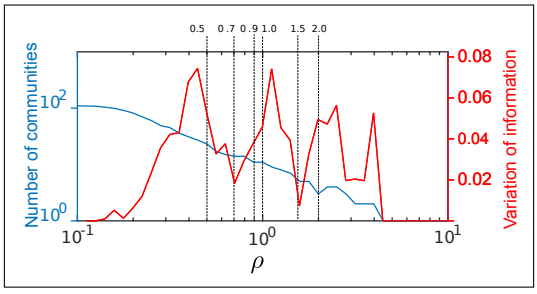


Figure 2.5. Stability method: Variation of information and number of communities detected by varying ρ.

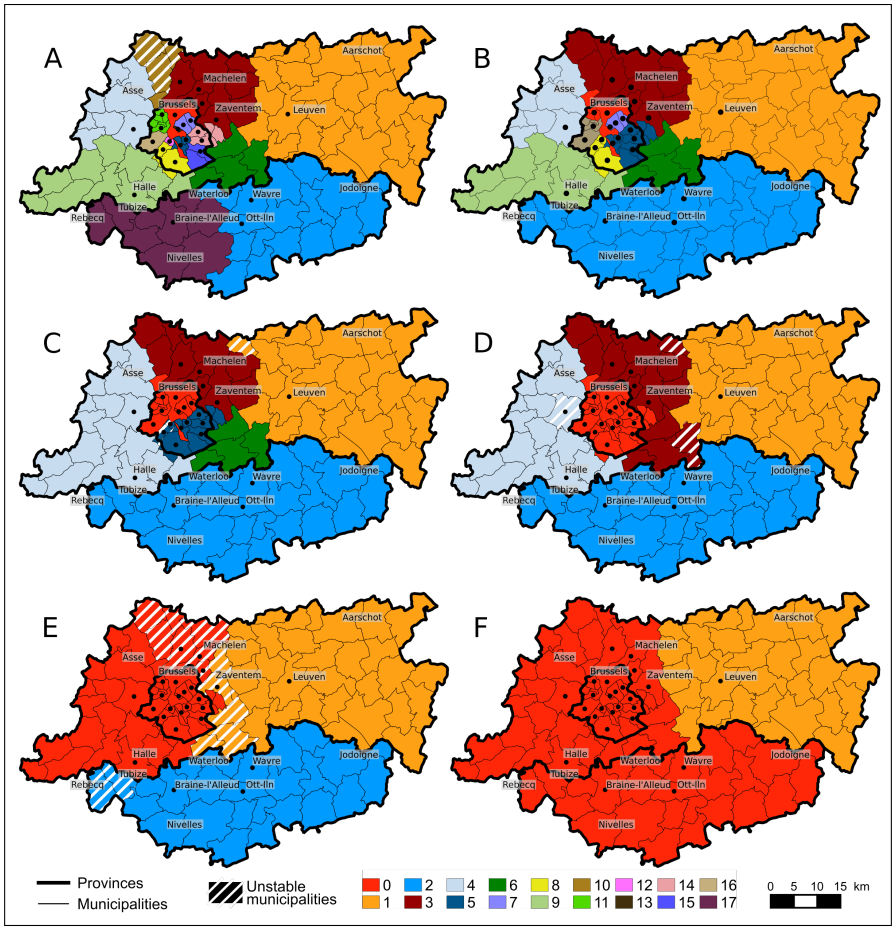


Figure 2.6. Partitions detected within the former province of Brabant for different value of ρ: A $= 0.5$; B $= 0.7$; C $= 0.9$; D $= 1$; E $= 1.5$; F $= 2$.

As mentioned above, we run 1,000 times the Louvain Method for the selected values of ρ to obtain $1,000$ partitions. We find that the number of communities detected for each ρ-value is not always the same for each run of the Louvain Method. For each value of ρ, Fig. 2.7 shows (a) the number of communities detected and (b) the number of times that these communities are found. When $\rho = 0.5$, the Louvain Method detects 789 partitions composed of 18 communities, 186 partitions of 17 communities and 25 partitions of 16 communities. Hence, for $\rho = 0.5$, the Louvain Method detects 18 communities that we consider to be the dominant community structure. Fig. 2.7 shows that each ρ-value leads to a very dominant community structure (with an occurrence of 80% or more) except for $\rho = 0.9$ where the occurrence of the most represented community is only 65%.

Next, for each of the 1,000 partitions detected with the selected values of ρ, we renumber the ID of the communities (for details see the Appendix) and analyze the dominant community structure. Fig. 2.7 shows that the number of communities detected by the Louvain algorithm is equal to 18, 11, 7, 5, 3, and 2 with $\rho = 0.5$, 0.7, 0.9,1.0, 1.5, and 2.0, respectively. As described in Section 2.2.1, when ρ increases, the partitions change from a large number of small communities to a small number of large ones.

Moreover, we find that each value of ρ leads to a delineation of communities that is not hierarchical. Indeed, the allocation of municipalities to communities is carried out under the objective function given by the Eq. 2.4. Each partition is thus independent of the others detected with different values of ρ. Table 2.2 represents the similarities between pairs of partitions and shows that the partitions obtained with values of ρ higher than one are less similar between them. We explain these differences by the gap between the values chosen for ρ. The gaps are smaller for values below one and, thus, the objective function of the maximization modularity is slightly modified.

Table 2.2. NMI values between partitions resulting from different ρ-values.

ρ	**0.5**	**0.7**	**0.9**	**1.0**	**1.5**	**2**
0.5	1	0.90	0.84	0.79	0.63	0.43
0.7		1	0.90	0.87	0.66	0.45
0.9			1	0.90	0.72	0.50
1.0				1	0.74	0.52
1.5					1	0.58
2.0						1

As expected, the size of the communities (corresponding to the sum of all the weighted edges) tends to be homogeneous across the partitions. This homogeneity can lead to the problem of *resolution limit*. To test the veracity

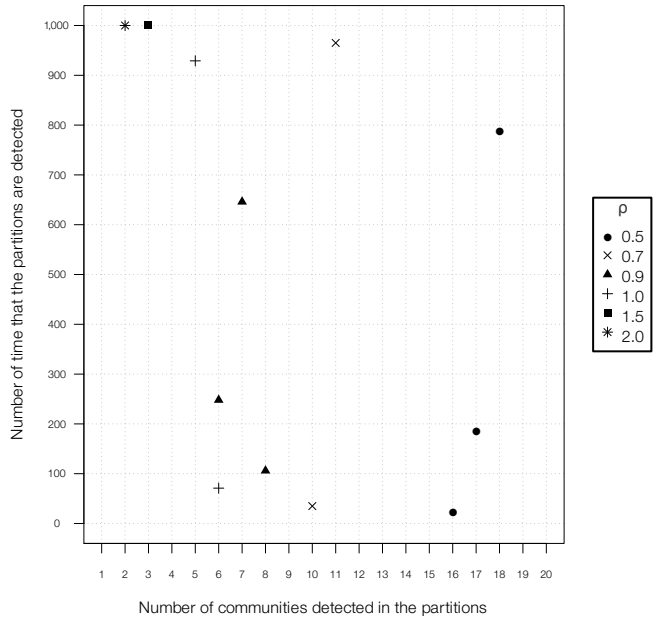


Figure 2.7. Number of times that each partition is detected for different values of ρ.

of the organization of the communities, we calculate the total weight of edges located within (internal weighted edges) and between (external weighted edges) communities for each partitions (see Fig. 2.8). Let us remind that the modularity finds the best partition under the *Cut* and *Diversity* criterions. The *Cut* (external edges) has to be minimized and the *Diversity* (number of equal size communities) maximized (Delvenne *et al.*, 2013). Here we look at how the internal and external weighted edges evolve with the increasing of ρ. With the rise of ρ, the proportion of internal weighted edges increases and when $\rho = 2.0$, 92% of the weighted edges are found within only two communities. When $\rho = 0.5$, we observe more external weighted edges located between communities than inside the structure and, for $\rho = 0.7$, we observe that 50% of the weighted edges are homogeneously distributed in and between the communities. Hence, it is hard to consider the partition obtained with $\rho = 0.5$ as a "good" community structure due to the large proportion of external weighted edges.

Next let us examine the spatial structures detected. Space is often polarized and jobs (commuters) concentrated in specific places. In the case of Brabant,

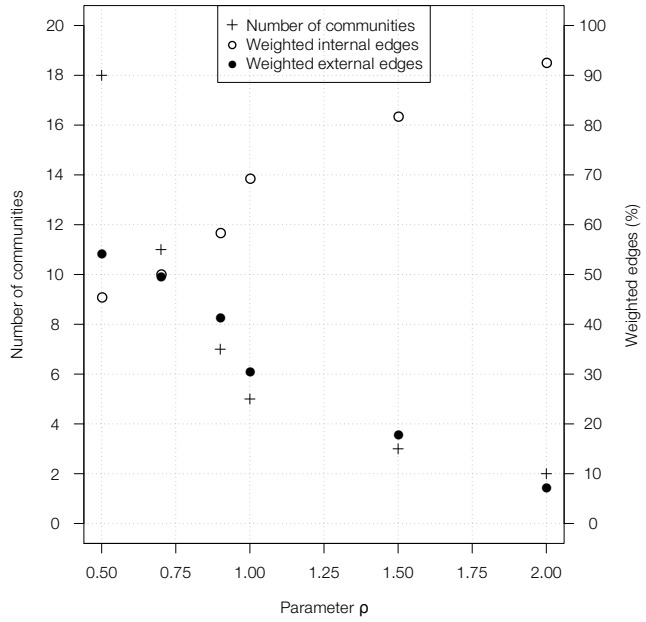


Figure 2.8. Number of communities detected, internal and external weighted edges for each partition.

these concentrations correspond to the highest values of weighted degree (see Fig. 2.4). When ρ is equal to 0.5, for example, the ten municipalities with the highest values of weighted degree are all classified in different communities that may be considered as "job basins" organized around some central places with high value of weighted degree.

The underlying geographical structure may play an important role in data analysis and, the observed communities are always sets of contiguous municipalities, although the Louvain Method does not take into account the location of the nodes. This confirms the strong effect of distance in commuting in Brabant (Riguelle *et al.*, 2007; Verhetsel *et al.*, 2010).

Within the province of Brabant, we detect two large communities that are slightly affected by variations in ρ. On the one hand, the community centred on Leuven (say Community 1) is characterized by a *compactness* value close to 0.7 in all cases, meaning that this community is self-centred (see Table 2.3). This corroborates the findings of Thomas *et al.* (2012). On the other hand, although the delineation of the southern community (Walloon Brabant) is little modified by an increasing ρ, the *compactness* values are close to 0.59.

Walloon Brabant has many internal commuting flows, and 40% of the flows are linked with municipalities located outside of the community. When $\rho = 2.0$ (Fig. 2.6.F), the community is merged with Brussels (Community 0), confirming the importance of the commuting flows between these two areas (Thisse and Thomas, 2010b).

Table 2.3. Compactness of the communities detected in the province of Brabant for different values of ρ.

Community		ρ					
ID	***Name***	***0.5***	***0.7***	***0.9***	***1.0***	***1.5***	***2.0***
0	Bruxelles	0.48	0.50	0.67	0.88	0.93	0.98
1	Leuven	0.71	0.71	0.71	0.71	0.70	0.70
2	Walloon Brabant (East)	0.56	0.59	0.59	0.59	0.59	
3	Vilvoorde	0.48	0.50	0.50	0.49		
4	Halle	0.42	0.42	0.45	0.45		
5	Ixelles	0.35	0.47	0.54			
6	Overijse	0.33	0.33	0.33			
7	Schaerbeek	0.29	0.33				
8	Uccle	0.37	0.40				
9	Dilbeek	0.38	0.38				
10	Meise	0.35					
11	Ganshoren	0.30					
12	Saint-Gilles	0.29					
13	Saint-Josse-ten-Noode	0.28					
14	Woluwe-Saint-Lambert	0.35					
15	Auderghem	0.30					
16	Anderlecht	0.32					
17	Walloon Brabant (West)	0.47					

Although the communities located in Brabant Wallon and in the BCR are merged together when $\rho = 2$, the linguistic border between the Flemish and the Walloon Regions is accurately followed by the delineation of the communities obtained with ρ equal to 0.5, 0.7, 0.9, 1.0 and 1.5. The linguistic border appears thus to be a barrier in commuting within the province of Brabant. Nevertheless, municipalities located within the BCR are connected with some Flemish places. For small values of ρ, three Flemish municipalities are included in the BCR (Linkebeek, Kraainem and Wezembeeck Oppem) and their number increases with ρ. This can easily be linked with the sprawl of the city (see Thomas *et al.*, 2012). When $\rho = 1.5$, the Western part of the area is totally merged with Brussels, confirming the history of commuting

(see Verhetsel *et al.*, 2010) and when $\rho = 2.0$ the community centred on Brussels crosses the linguistic border. This Brussels community presents a high value of compactness (0.98) revealing that the majority of the weighted edges are within the community. We notice that some municipalities squeezed between Brussels (Community 0) and Leuven (Community 1) have an unstable classification when $\rho = 1.0$ (Fig. 2.6.C). Both cities attract commuters and the Louvain Method is unable to provide a robust solution, leading to the the determination of fuzzy borders. This is confirmed by Fig. 2.6.E and Fig. 2.6.F. When $\rho = 1.5$, these municipalities belong to the Leuven community, while when $\rho = 2.0$, they are incorporated in the community around the BCR. This area is polarized by both cities in terms of residence and employment (see Vandermotten *et al.*, 2009a).

The advantage of introducing ρ in the Louvain Method is to modify the objective function that optimizes modularity and to detect partitions with communities of different sizes. ρ allows us to detect small communities within the BCR and to understand how the municipalities interact. Within the BCR, the number of communities detected is equal to nine when $\rho = 0.5$, two when $\rho = 0.9$ and one when $\rho = 1.0$. Each of these partitions can be easily linked with the literature on Brussels. The communities detected with $\rho = 0.5$ and 0.7 (Fig. 2.6.A and Fig. 2.6.B) are in line with the historical suburban organization of Brussels (see Eggerickx, 2013) where each municipality presents its own socio-economic characteristics resulting from its particular history. Moreover, when $\rho = 0.9$ (Fig. 2.6.C), the BCR is split in two parts: the north-western and the south-eastern area. The population in the south-eastern area has a higher average income. The patterns detected by the Louvain Method perfectly fit with the socio-economic reality as depicted by Vandermotten *et al.* (2009a). The small central communities detected with ρ equal to 0.5 and 0.7 are characterized by rather small *compactness* indices (0.25 to 0.48) meaning that these communities are more connected with the rest of the study area. Consequently, when $\rho = 0.9$, these many small structures merge into two larger communities (north-west and south-east) with *compactness* values equal to 0.67 and 0.54 respectively (see Fig. 2.6.C). When $\rho = 1$, we observe only one central community that almost perfectly corresponds to the BCR (Fig. 2.6.D). The *compactness* value of this community is 0.88. The small communities located within the BCR are hence more affected by the variations of ρ than the communities having larger values of *compactness*. As a general overview, the steps illustrated in Fig. 2.6.A to 2.6.F corroborate the history of urban sprawl around Brussels as well as the history of commuting and transport infrastructure, all of them constrained by administrative realities (see Dujardin *et al.*, 2007; Poelmans and Van Rompaey, 2009).

2.3.2 Effect of self-loops

The effect of the diagonal in origin - destination matrices (self-loops) has already widely been discussed in geography (Grasland, 1999). One can show (see Delvenne *et al.*, 2013; Lambiotte *et al.*, 2014) that adding self-loops to each node of a weight proportional to the degree of the node is strictly equivalent to changing the resolution parameter, when identifying communities. Indeed, the self-loops change the *Cut* term of Q_ρ by scaling the total weight $2m$, but leave other terms unchanged. This is congruent with a similar observation of Arenas *et al.* (2008) who deals with constant-size self-loops instead. A second set of sensitivity analyses is thus performed in order to quantify the influence of self-loops on the results. For the sake of clarity, we present only comparisons between partitions with the same number of communities (11, 8 and 4). Fig. 2.9 illustrates the case of eight communities. To obtain eight communities by including or not self-loops, the Louvain Method was applied with $\rho = 0.9$ for the network including self-loops and with $\rho = 0.7$ when not. As expected, smaller self-loops tend to produce larger communities and we were thus led to decrease the value of ρ to detect eight communities.

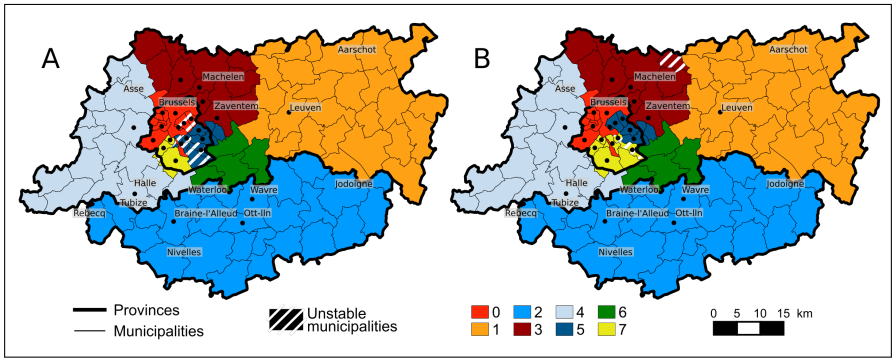


Figure 2.9. Partitions detected in the former province of Brabant (A) with and (B) without self-loops (eight communities).

Results show that self-loops are sufficiently congruent in size with the total degree of communities, to leaves partitions unchanged. Nevertheless, to obtain the same number of communities, the resolution parameter has to be rescaled. Deleting/including self-loops has little effect on the composition and the delineation of the communities evidenced by the NMI values that range between 0.98 and 0.93. Fig. 2.9 shows that the few differences are mainly located within or close to the BCR (Communities 0, 4, 5 and 7); the delineation

and the composition of these communities are slightly modified by the removal of the self-loops. Let us note that although Community 5 has the same number of municipalities (Table 2.4), its composition and delineation are different. The *compactness* slightly decreases for Communities 0 and 5 when deleting self-loops and, slightly increases for Communities 4 and 7 (Table 2.4). These variations are small and do not lead to important changes in the connections of these communities with the others. Let us here recall that $1,000$ runs were computed for each use of the Louvain Method and that the municipalities that are not always classified in their dominant structure are hatched on Fig. 2.6. Hatched municipalities also correspond to those that are affected by the self-loops; internal links slightly affect the results of the Louvain Method.

Similarly, we extended the analysis to the entire country (see also Section 2.3.3). Fig. 2.13.A refers to the partition of the 589 municipalities into ten communities including self-loops, while Fig. 2.13.B excludes self-loops. Both partitions are once again very similar visually, which is also evidenced in terms of NMI (0.916). The main difference between the two partitions is the extension of the community centred on Brussels, which is much larger in Fig. 2.13.A than in Fig. 2.13.B. When excluding self-loops (Fig. 2.13.B), the eastern part of Brabant Wallon is not included anymore in the community centred on Brussels. Interestingly enough, this classification is unstable as illustrated by the hatching. The Louvain Method is not very sensitive to the self-loop at any size of the study area for commuting movements. However, it would be interesting to test the sensitivity on another individual dataset as well. Indeed, we used the aggregated commuting flows between municipalities and the importance of self-loops is here related to the values of weighted degree. From a geographical point of view, it is expected that a municipality that concentrates a lot of commuters (and thus can be called a *job centre*) also has a large self-loop too. In consequence, the small sensitivity of the Louvain Method to self-loops is clearly not due to the method itself, but to the strong structure of the data.

2.3.3 Delineation of the study area

The last set of sensitivity analyses deals with the extension of the study area. If we limit the analysis to the origin-destination matrix of flows between the 111 municipalities of the province of Brabant (Fig. 2.10.B), do we get the same results as if we consider all the flows between the 589 municipalities of Belgium and zoom in on our study area (Fig. 2.10.A). If we ignore the relationships between the municipalities of Brabant and the rest of the country, do we get the same partitioning? Are the other cities exerting an influence? Geographers are familiar with this problem for several kinds of quantitative models (Horner and Murray, 2002; Jones *et al.*, 2016) even if often ignored (Jones *et al.*, 2017). In this study more than others, the urban network is tight with a high density

Table 2.4. *Compactness* values of the communities detected with the inclusion of or without self-loops (the number of municipalities within communities is given in brackets).

Community ID	With self-loops		Without self-loops	
0	0.670	(11)	0.603	(8)
1	0.711	(26)	0.711	(26)
2	0.587	(27)	0.587	(27)
3	0.501	(12)	0.501	(12)
4	0.446	(18)	0.449	(19)
5	0.466	(8)	0.407	(8)
6	0.328	(4)	0.328	(4)
7	0.402	(5)	0.495	(7)

of closely located regional centres.

For the sake of clarity, we limit ourselves to two illustrations and for the same number of communities within the province of Brabant (see Fig. 2.10). The NMI between these two maps takes the value of 0.86, hence confirming the visual observation. Both maps look alike without being perfectly identical. The same partition is observed with the exception of (a) the western peri-urban area where the urban network is tight (high density of closely located regional centres: high values of weighted degree), tending to a strong historical commuting movement (Verhetsel *et al.*, 2010) and (b) with the exception of a small community within the BCR that appears in Fig. 2.10.A and not in Fig. 2.10.B. These exceptions correspond to the municipalities of Saint-Josse-ten-Noode, Evere and Schaerbeek which tend in practice to federate several public services within the urban agglomeration (an unique Police Zone 5344). Interestingly, in the partitions illustrated in Fig. 2.10.A there are more unstable classifications. When eliminating the flows with all other Belgian municipalities, the network is smaller and simpler and could lead to more robust results since the Louvain Method has fewer partitions to explore than in the larger network (see Fig. 2.10.B).

2.3.4 Infomap

This section aims to compare the partitions detected by the Louvain Method with those obtained by the Infomap algorithm. We propose a slight improvement and refer to Rosvall and Bergstrom (2008) for further methodological development. Infomap is an algorithm based on the idea that a random walker, following each weighted edges, should spend most of its time inside communities and rarely travel across communities. To compute the steps of this random walker, the authors represent the movements at two

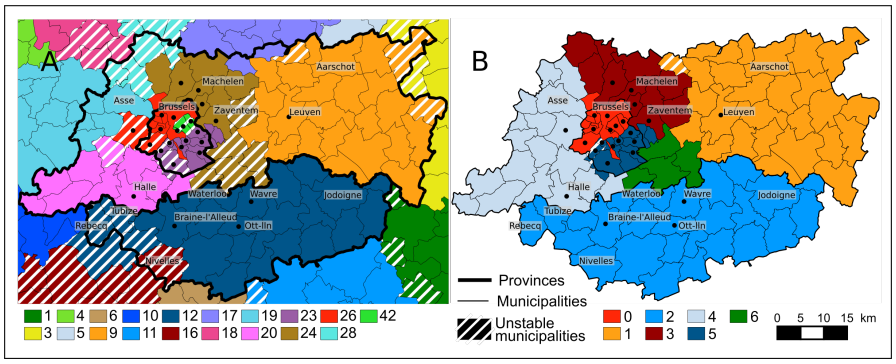


Figure 2.10. Communities detected in Belgium and zoomed on the former province of Brabant (A), and detected in the former province of Brabant (B).

levels. The first level encodes the different communities and the second level the individual nodes. Thus they look for a partition of community c made of N nodes that minimize the expected description length of a random walk. The Infomap version used here is the standard version. It is hence impossible to detect communities at different levels of resolution.

Let us compare the partitions obtained by both methods (Louvain and Infomap) for a given number of communities. For Belgian commuting movements, Infomap detects 14 homogeneous communities. Based on the methodology described in Section 2.3.1, the value of $\rho = 0.6$ allows Louvain to detect 14 communities as well. The two partitions of Belgium into 14 communities are similar (NMI=0.82). Results are mapped and analyzed at the scale of the former province of Brabant (see Fig. 2.11). The former province of Brabant is dominated by three major communities by the Louvain Method but only two by applying Infomap. The community centred on the area of Leuven (Community 2) is detected by both algorithms and the composition and delineation tend to be similar (30 municipalities by using the Louvain Method and 27 municipalities by using Infomap). The community located in the area of the BCR (Community 1) is detected by both algorithms but with two major differences. On the one hand, this community is much larger in the Infomap partition (101 municipalities instead of 57 with the Louvain Method), and, on the other hand, the majority of the municipalities of Walloon Brabant are classified to the BCR by applying Infomap, although these are allocated to another group in the case of the Louvain Method (Community 9: Namur-Wavre). An interesting point is the instability in the classification of some municipalities with the Louvain Method (see Fig. 2.11), such as the municipalities located within the province of Walloon Brabant

and the one squeezed between one and two (BCR and Leuven).

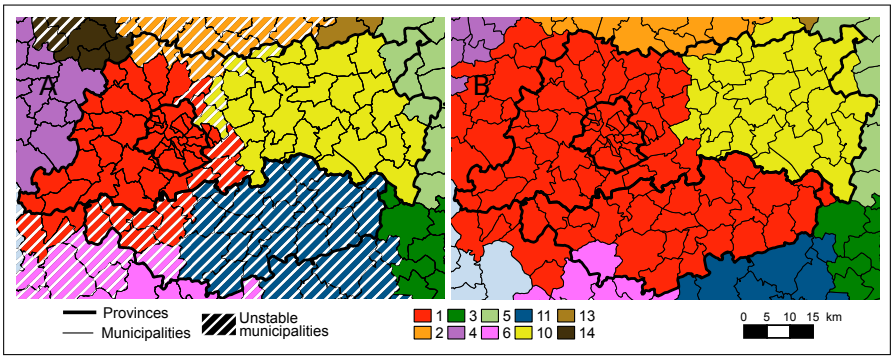


Figure 2.11. Communities detected by the Louvain Method (A) and by the Infomap algorithm (B), computed for Belgium but presented for the former province of Brabant.

2.4 Conclusion

Optimizing modularity using the Louvain Method is a quantitative approach that enables to analyze the organization of a network that can be applied in the field of geography. The method partitions space into communities of highly connected places in terms of commuting flows. We analyzed the robustness of the results with respect to 3 modifications. Changing the parameter ρ has the most important impact while the inclusion or not of self loops (intra-communal commuters) and the definition of the study area have respectively no or little impact. The choice of ρ is essential but there is no unique methodological guideline for choosing it and each value of ρ leads to a different partition. This is found to be a major shortcoming of the approach. We used the method developed by Delmotte *et al.* (2011) and implemented in Delvenne *et al.* (2013) that quantifies the stability of partitions depending on ρ to detect the level of organization of commuting movements as well as the robustness of the partitions detected at different levels. Moreover, we examined the partitions in detail and tested how the communities are structured in terms of internal and external flows to characterize each group. Further research should be directed towards analysing to find whether the most stable partitions are also the best partitioning of space or not. Interestingly enough, the impact of the delineation of the study area corresponds to a problem familiar to geographers: by modifying the size of the study area, we ignore a certain amount of information that may originate

outside of the delineation and characterise different realities. The small differences observed with/without self-loops correspond to places that are not classified unequivocally in the 1,000 runs. Such places are mainly located at the outskirts of the communities.

The Louvain Method is an interesting method for extracting geographical information from relational data. The method delineates communities within networks and we approximate their centres by calculating the weighted degree of each municipality. In our case, job basins are spatially continuous but without relying on locational information. Space matters, and geographical maps can easily be interpreted in terms of the socio-economic geography of Belgium. Not surprisingly, in the former province of Brabant, the linguistic and regional borders are not permeable (at some exceptions located close to Brussels). Moreover, two communities are particularly little affected by the sensitivity analyses: Leuven and the Walloon Brabant. We also show that unstable municipalities are squeezed between Brussels and smaller secondary cities, especially in the eastern and western areas. This strong effect of space is also present in case of the second community detection algorithm used for comparative reasons (Infomap). Partitions are similar to those obtained with the Louvain Method but with some exceptions. Municipalities located close to the borders of the communities (job basins) show instability features. Space, and the networks of cities within their environments have a strong effect on commuting patterns: nearby places are more intensively linked than distant places. However, urban realities are by nature complex, and our study is limited only to one aspect (job basins). With the next Chapters, we propose to compare these results with other kinds of flows and link them to the socio-economic, environmental or other characteristics of the nodes: analyses based on migration data (Chapter 4), train schedules requests (Chapter 5), mobile phone calls combined with socio-economic variables (Chapter 7) and the real-time tracking of trucks (Chapter 6). Future research in this direction might help planners to improve their perspective on the future of cities and particularly that of Brussels.

Appendix

A.1 Extraction of the dominant community

With the multiple runs of the Louvain Method, the nodes are characterized by a list of 1,000 identification number (id) corresponding to the community number. It is important to underline that the allocation of nodes to communities is specific to each run of the Louvain Method, implying that the ID of the communities is renumbered at each run. Hence, a community numbered “one”

in the first run is not automatically labelled "one" in the other simulations; so we have to make the ID of the communities comparable through the runs of the Louvain Method. The methodology is illustrated in Fig. 2.12.

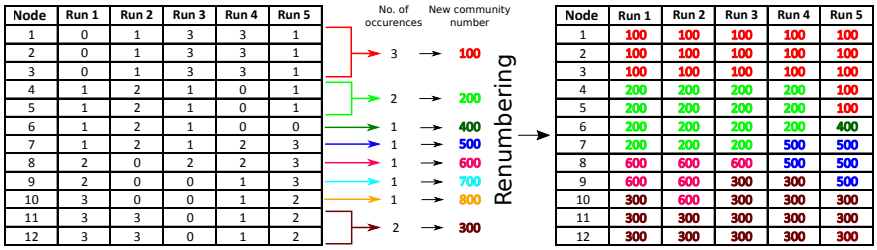


Figure 2.12. Renumbering of the communities for five runs of the Louvain Method.

First, each node is characterized by a list of 1,000 elements corresponding to the communities where the nodes are allocated. We call this list the *signature of the node* and calculate the number of occurrences of each signature in the results. Second, the list of signatures is ordered by descending order of occurrence and we attribute to each a specific id that we call the *new community number*. Third, the highly represented signature is used to renumber the communities detected. The renumbering starts with the nodes having the highest represented signature. The initial community numbers are modified with the *new community number*. Moreover, this renumbering is applied for the entire run and so, if the initial community number is still present, it is renumbered with the *new community number* too. This special case is illustrated in Fig. 2.12. In run 5, the nodes with the highest represented signature are nodes 1, 2 and 3 which belong to Community 1. In this run we observe that a number of other nodes are initially classified within this community 1 (nodes 4 and 5). So, all, these nodes (1, 2, 3, 4 and 5) are renumbered with the *new community number* 100. Once all the initial IDs of the communities have been modified by the *new community number*, the same procedure is applied to the second most frequent signature. This renumbering is applied until all the communities are renumbered. Finally, the dominant community (most frequent community) is extracted for each node and the robustness of the nodes is calculated.

A.2 Belgian partitions with(out) self-loops

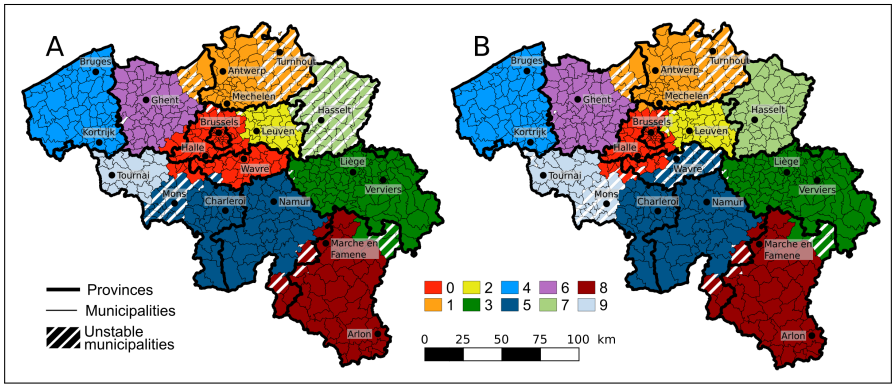


Figure 2.13. Partitions detected in Belgium (A) with and (B) without self-loops (10 communities).

3

Residential basins in Belgium: two methods, one reality?

The Louvain Method provides a new way for geographers to partition a space my means of interaction fluxes. This Chapter compares the Louvain Method to the Anabel method that was initially developed and used by the INSEE since 1980 for grouping places by means on their interrelationships. The main objective here is to compare the advantages and limitations of both methods, while providing a critical geographical understanding of the spatial results. Two databases are used: a synthetical data base (theoretical) and a real world data set (inter-communal migrations in Belgium).

Outcomes remind us that partitioning space is not a trivial exercise, that the results strongly depend on the objectives, on the methods and thresholds used, and that policy decisions should always be taken in a cautious and informed manner.

This Chapter is based on works made in collaboration with the *Institut wallon de l'évaluation, de la prospective et de la statistique* and already published in the *Espace géographique* (see Adam *et al.*, 2018a)

3.1 Introduction

Partitioning of a territory by means of data exchange matrices has been common practice in geography and regional science for decades. Partitioning means identifying groups of neighbouring places having more relationships between them than with the rest of the territory. Following the development of big-data, newly developed methods are commonly applied in the clustering of places in origin-destination matrices. These latest methods advantageously provide prompt analysis for overwhelming amounts of data retrieved by information and communication technologies and other big-data (Fortunato, 2009), and thereby efficiently detect communities of places with privileged exchanges.

These methods, being traditional or not, all aim at synthesizing flows and providing maps of territorial relationships, but it is often difficult to compare them because firstly, the flows used can vary in nature and frequently occur at different geographic scales, secondly, the objectives of the developers of these methods diverge, and lastly, the criteria and methodologies used are based on different developments which remain poorly documented.

Here the proposed exercise aims to elucidate the contributions from a traditional method for spatial partitioning into basins (Anabel) and from a recent method of community detection (Louvain Method - Chapter 2), both of which will be applied to the same database (changes of residence in Belgium). The Anabel method is widely used by INSEE, from 1998 to 2015, and implemented by the Walloon Institute for Evaluation of Forecasting and Statistics (IWEPS) to partition the Walloon Region into employment basins (de Wasseige *et al.*, 2000) and residential basins (Charlier *et al.*, 2016)[1]. The Louvain Method detects communities of places with privileged exchanges by means of modularity (Blondel *et al.*, 2008). Today, this method is extensively used for detecting, inter alia, employment basins (Montis *et al.*, 2013) or telephone areas (Blondel *et al.*, 2010). The interest of this paper is thus to compare the two methods, both from a methodical perspective and in the interpretation of the results, in order to offer recommendations to users and readers.

The Anabel method is briefly presented in Section 3.2 while the Louvain Method was already accurately described in Chapter 2. Both methods are then applied to a theoretical data set before being applied to the migration exchanges between Belgian municipalities. The aim here is to highlight the respective contributions of these two methods for a better understanding of the spatial planning and for providing partitions that are more suited within the political framework of territorial development, housing or mobility.

[1] Other methods have also been used in Wallonia, mainly to define school areas (Delvaux, 2005).

3.2 The Anabel method

Anabel (location pair analysis for local research) is a software tool developed by INSEE, for gathering observations based on exchange flows (Théré and Séguy, 1998). The tool (initially known as Mirabelle then MIRABEL) has been used since the 1980s, to develop different forms of spatial partitioning, firstly applied to employment basins and thereafter to metropolitan areas and residential basins (Théré and Séguy, 1998). This tool makes it possible to study and to synthesize the relationships forged between territorial entities and to highlight the various hierarchies (Théré and Séguy, 1998).

Based on an origin-destination matrix, the basic principle is to group spatial units according to the importance of their connections computed from absolute flows[2]. The classification process is performed iteratively. In the starting matrix, the algorithm looks for the connection of maximum value, then it aggregates the two entities selected into a new entity. Once merged, the two starting nodes (entities) cease to exist, but are represented by this new composite entity. The connections between this new entity and all the remaining entities are then recalculated and the process continues the search for a new connection of maximum value. The generated entities become progressively larger and aggregate. The process may be simplified in the form of a tree: the roots of this tree, representing each of the entities to classify gradually converge on the basis of maximal connection, clustering an increasing number of entities (Théré and Séguy, 1998).

A stopping criterion is applied prior to the final stage (a single basin). Several options have been used. Accordingly, it is possible to prohibit a preformed entity to absorb another entity whenever it responds to suitable size criteria (number of municipalities, population, etc.) or if the connection which unites them is weak (in this case the aggregation process cuts off and certain entities remain unassigned to a basin). It is equally possible to prevent two preformed basins from merging with each other. Defining the criteria and thresholds to be considered depends on the objectives of the study. Synthetically, during the iterative process the entities merge in a known and transparent sequence. When the iterative process is interrupted, the position is therefore fixed at the number of entities and their size. The order of the aggregation is important as it is definitive and will influence the size of the entities formed in comparison to other entities and thus, the classification of entities in the final result. A rebound effect may appear: "*a geographical unit may be pulled into a cluster with a stronger connection than the geographical unit previously joined to it, quite simply because it is pulled in by both the cluster and the previously aggregated unit*" (Théré and Séguy, 1998).

[2]This connection value may be calculated in several ways depending on the type of flow (including a single flow direction or not) and according to the target.

This method may amount to a Hierarchical Ascending Clustering (HAC), although here the notion of *connection* between spatial entities substitutes the notion of *distance* (Théré and Séguy, 1998). The choice of the stopping criterion for the aggregative process may be made in studying the initial tree representation. While cut-off thresholds are determined a priori, a certain amount of subjectivity is introduced into the results.

3.3 Application of the two methods to a theoretical network

3.3.1 The synthetic network

Let us consider a theoretical network of 44 entities (squares) among which migrations are observed and illustrated by the arrows between the centres of the squared entities. Migrations are distributed arbitrarily in order to highlight three main urban poles: poles 1 and 2 are separated by an intermediary zone where the entities exhibit no clear pattern in the exchanges, while the third pole is more isolated and less connected. Figure 3.1 illustrates the organization of the absolute migration flows between the different entities constituting this arbitrary and theoretical network. The information on each entity is summarized at its centre. For the sake of readability, the connections (the sum of flows between two entities) are ranked and movements under 100 are not shown; these small exchanges are homogeneously spread over all entities.

3.3.2 Application of Anabel

Using the Anabel method, the matrix of absolute migration flows is generally transformed into a matrix of relative connections. Depending on the objective set, several different formulas exist to calculate a connection between entities. The aim here is to highlight the residential migration exchanges between entities, both inward and outward, and not the attractiveness of one over the other. Indeed, the centrifugal migration flows (from the centre to the periphery) can be higher than the centripetal flows, yet it is not the attractiveness of the centre that we seek to highlight, but all relationships between entities. The linkages between entities A and B are thereby determined as follows:

- Linkages of A in comparison to B: $(flow_{(A,B)} + flow_{(B,A)})/(inward_{(A)} + outward_{(A)})$

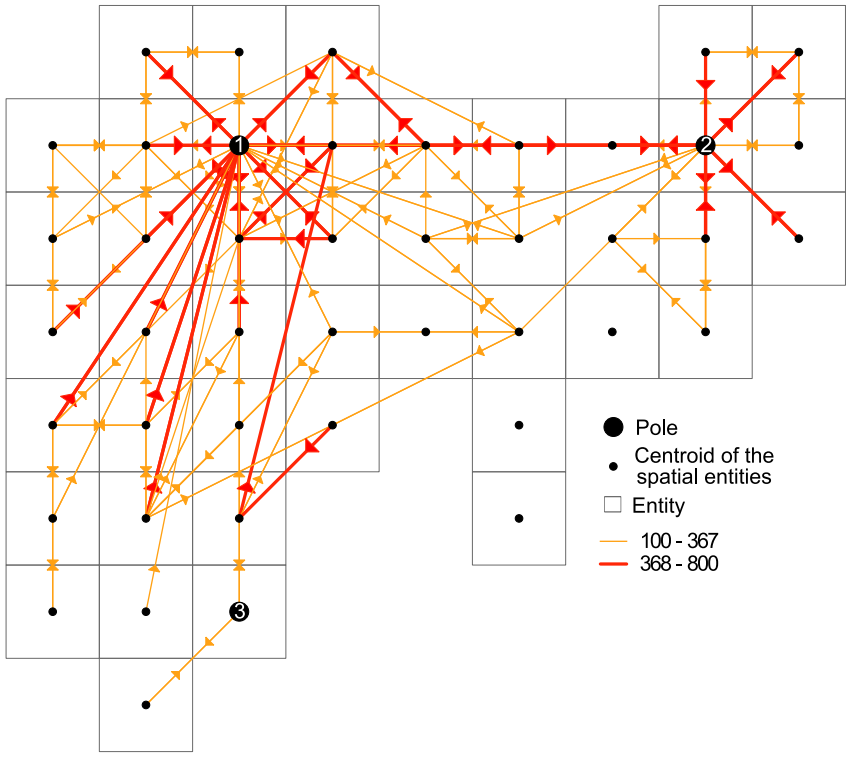


Figure 3.1. Theoretical network of 44 entities.

- Linkages of B in comparison to A: $(flow_{(B,A)} + flow_{(A,B)})/(inward_{(B)} + outward_{(B)})$

The value of the AB cell therefore corresponds to the share of the flows between A and B within the totality of the flows (inward + outward) of the entity A along with all the entities. This AB connection thus reflects the significance of the connections of A in comparison to B. This relative approach allows for the consideration of the size of the flow through entities, independently of the weight of each entity; namely, the total number of inhabitants or the overall size of inward and outward flows.

The aggregative process is then initiated as defined in the method outlined above. A first basin starts to form around pole 2 (red) where connection exchanges are the highest. Next, basins are formed around pole 3 (green) and pole 1 (blue). The basins surrounding poles 1 and 2 expand until the

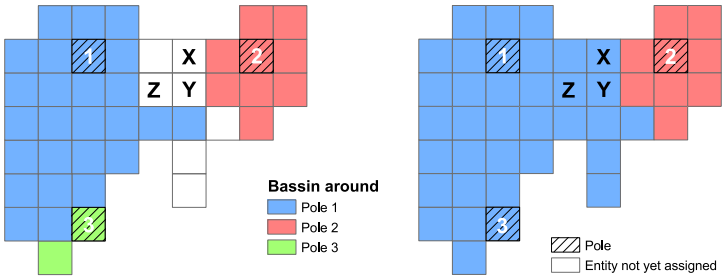


Figure 3.2. Anabel method applied to the theoretical network in iteration 32 (high) and 42 (low).

33 iteration, basin 1 absorbs basin 3[3]. Figure 3.2.A shows the situation just before this absorption. Thereafter, only the basin linked to pole 1 continues to increase in size until all the entities of basin 3 (green) aggregate with basin 1 (3.2.B – iteration 42).

3.3.3 Application of the Louvain Method

The Louvain Method is first applied on absolute exchange values (a *native* application of the method) and then – for comparison purposes – on the relative values used in the Anabel application by the Walloon Institute of Studies, Prospective and Statistic (IWEPS).

Two values of the parameter ρ are here illustrated (1 and 2). These two values make the Louvain Method results more easily comparable to the results of the Anabel method. When $\rho = 1$, three homogeneous sized communities quite are detected, and when $\rho = 2$, two communities emerge (Figure 3.3). The Louvain Method seeks to maximize a modularity target specific to the value of ρ. Henceforth, the cells of community 3 are predominantly assigned to community 1. The Louvain Method is then applied to data expressed in relative terms. In applying two different values of ρ (1.25 and 2), 3 and 2 communities are respectively identified by Louvain (Figure 3.3). The composition and delimitation of these two communities is exactly identical to the communities detected by the Louvain Method in the absolute exchanges, which shows that the structures detected remain strong and stable whatever migratory matrix is used (relative or absolute).

[3]The merging together of basins is optional and configurable in Anabel.

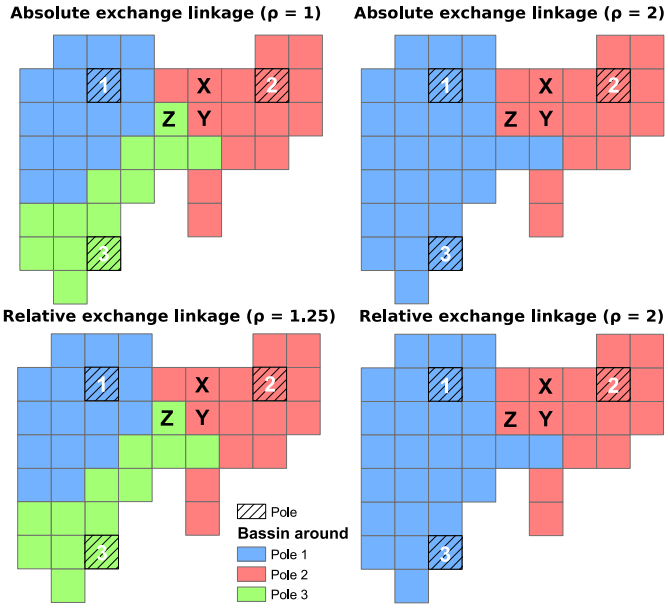


Figure 3.3. Louvain Method applied to the theoretical network.

3.3.4 Method comparaisons

The two methods lead to the partitioning of the 44 entities into two or three groups, depending on the timing chosen to break the iteration-aggregation process of Anabel[4], and based on the value of ρ introduced in the Louvain Method.

Once the 44 squared entities are divided into three groups, the Anabel method and the Louvain Method display basins centred on the three poles previously defined, but with different compositions/boundaries. Table 3.1 summarizes the number of entities classified in each group for the two methods. Anabel detects basins of various sizes, whereas the Louvain Method tends to make the size of the communities more homogeneous. In the Anabel method, the classification of the nodes occurs throughout the process. When we let the process run, the basin around pole 3 merges with the basin around pole 1; this latter becomes increasingly *greedy*, its weight within the matrix growing progressively heavier. The basin of pole 1 eventually absorbs the remaining entities, notably through the rebound effect. This latter can be explained as follows: although entities X, Y and Z (Figure 3.2.A and .B) initially had

[4]Where the aggregation of basins is not authorized (tool option Anabel), three different basins are detected, even at the end of the iterative process.

Table 3.1. Number of entities composing each group (basin/community), according to the method used and the number of groups selected (2 or 3).

	Number of groups	Group 1	Group 2	Group 3
Anabel	3	26	9	2
Anabel	2	35	9	-
Louvain	3	16	15	16
Louvain	2	28	16	-

linkages with pole 2 superior to pole 1, they end up in the basin of pole 1, as just prior to their merger, they maintain a larger share of exchanges with the entire preformed basin of pole 1 instead of with the basin of pole 2. This effect commonly concerns entities located in a zone of shared influence. This classification is not the same in the Louvain Method results (Figure 3.3), the entities in the area of shared influence between poles 1 and 2 are classified in the community incorporating pole 2.

In classifying the entities into two groups, the basin of pole 3 is included in the group of pole 1, regardless of the method. The basin of pole 1 is, in either case, the larger basin (35 entities for Anabel and 28 for Louvain).

An initial observation stands out: applied to the same data set the two methods proceed differently in spatial partitioning. With the Anabel method, entities merge solely on the basis of maximal linkages and these mergers occur according to a hierarchy of observed and recalculated connections after each aggregation. The Louvain Method, however, considers all the exchanges and seeks a compromise between the maximum number of communities (*diversity*) and the minimum number of connections between communities (*cut*). This methodological difference is illustrated in a simplified way in Figure 3.4. That is, locations already classified into two groups (red and blue dots) and a black dot to be classified. Anabel puts this black dot with the group of blue dots as the exchange (black line – value 45) is the strongest[5]. The Louvain Method recognizes all the exchanges related to this black dot. The value of the connection with the blue dots is clearly 45 and the strongest, but the sum of the other exchanges around the red dots is 55, in other words a higher value. The black dot will thereby be classified with the red ones using the Louvain Method.

In conclusion, the partitions obtained with the Anabel method are the result of aggregating entities according to the hierarchical affinity tree, which invariably involves treating the strongest connection. The static aspect of this tree favours the development of partitions around the spatial entities that are the most closely connected. Thereafter, the process stimulates the development

[5]Unless all the red dots already form a basin in the aggregation process.

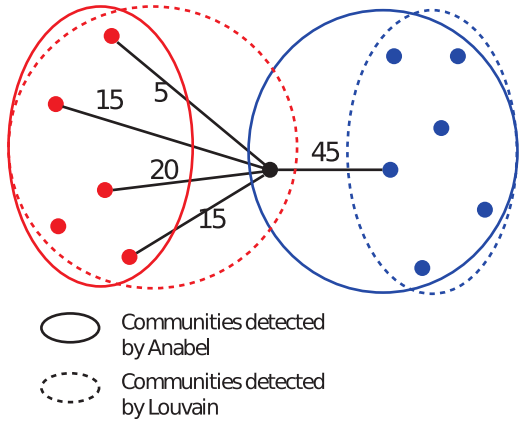


Figure 3.4. Comparison of the results for the methods Anabel and Louvain on an imaginary data set.

of these nuclei notably through the rebound effect. Throughout the aggregative process, Anabel serves to highlight evolving hierarchies between polarities and their area of migratory influence within the territory, although the final result is contingent upon the priority limits established for the model.

The Louvain Method, however, is global and considers all the connections of the network when clustering nodes into communities. In this way, communities are composed of nodes which have intensive exchanges between them. As this method is non-hierarchical, there is no classification tree. Therefore, even though Louvain detects groups of nodes, we cannot name these structures basins since a basin requires a centre. Let us recall that in Anabel, the centre of the basins is established during the iterations. In the Louvain Method, the polarities may potentially be defined a posteriori by calculating the node degrees; those exhibiting high values subsequently indicate that these nodes concentrate a larger proportion of flows (Adam *et al.*, 2018b). The community partitions correspond to a balanced mathematical optimum seeking to maximize modularity. By consequence, it requires the creation of roughly similar sized communities (as regards to the amount of flows) given that a basin centred on a cell could have a greater (or lesser) migratory influence on the territory than another.

3.4 Application of the two methods to Belgium

3.4.1 Data matrix

The Belgian National Register constitutes an authentic and legitimate source of data on natural persons. It centralizes information from the population register (Belgian and foreign population: foreign residents in Belgium, allowed to settle or stay longer than 3 months), the waiting register for applicants of refugee status, the diplomatic and consular registers and also from the registers of international civil servants (European Union, North Atlantic Treaty Organization, etc.). It is the basis of Belgium's demographic statistics. Data recording is assigned to the municipalities and to the Immigration Office for the waiting register.

The relational data matrix used here is therefore the one concerned with changes of residence, that is to say, the population movements (migrations) between the 589 Belgian municipalities, registered from January 1st 2012 to 31 December 2014 (it is important to notice that movements made by the foreign population is not considered here). This data clearly has its limitations (Eggerickx *et al.*, 2010) as some inhabitants may move several times during the period studied, while others move within the same municipality, and ultimately, intra-municipal moves are particularly more frequent as the municipality is populated. In the context of this study, migrations between an address in Belgium and another country, are excluded because the National Register does not specify the foreign address (it only states the foreign country).

First, let us point out that the 19 municipalities constituting the BCR have been aggregated to form a single entity. Although this aggregation derogates from the chosen study at municipal level, it avoids highlighting sub-regional migration influences in Brussels, as well as positioning Brussels in relation to other cities primarily constituted by one large municipality (such as Antwerp or Ghent). Following this aggregation for Brussels, migration exchanges will be studied within a 571 x 571 matrix. Secondly, the comparison exercise performed here uses inter-communal migration data and therefore does not take into consideration moves within a given municipality. With the Louvain Method, the intra-municipal migrations may be considered, but are not used here to facilitate the comparisons between the methods.

3.4.2 Anabel method Results

In order to avoid a final partitioning retaining undersized local basins, the basins are authorized to merge together during the process (Charlier *et al.*, 2016). It is equally desired that each municipality is assigned to a basin. In order to avoid ending up with only one basin, it is important to establish a stopping criterion in the aggregative process: here the cut-off is set to the

moment when all the municipalities are assigned to a basin. This timing coincides with the partitioning of Belgium into 8 basins (Figure 3.5) and occurs after 562 iterations. Mapping the aggregative process at various times (Figures 3.6.A and 3.6.B) makes it possible to view and study the way in which this final partitioning is generated.



Figure 3.5. Residential basins based on the Anabel method – authorized inter cluster aggregation situation after 562 iterations.

After 290 iterations (Figure 3.6.A), basins are formed around the principal centres (poles) of migration exchanges (Charlier *et al.*, 2016): the BCR, Antwerp, Ghent, Bruges, Tournai, Charleroi, Namur and Liège. Smaller basins have also begun to emerge (Oostende, Mouscron, Chimay, Verviers, Eupen, Sankt-Vith, Bastogne, Arlon, Virton). Most of them will then be absorbed, showing the local significance of these polarities (Figure 3.5). After 536 iterations (Figure 3.6.B), three large basins cover the Region of Flanders while in Wallonia, numerous small basins coexist in addition to places not yet assigned, namely, the west of the province of Luxembourg and the southeast of the province of Namur. The lack of polarity driving migrations in this zone results in migration exchanges of a more diffuse nature, which allows the development of small areas around minor polarities (Ciney, Bastogne, Virton, Aubange, etc.). The basins of Liège and Namur are going to gradually absorb the municipalities of this zone.

Extending the aggregation process until all the municipalities are included in a basin (Figure 3.5) means that some basins aggregate well before, due to the higher rate of exchange. All of the Belgian territory is covered in large basins encompassing small-scale basins, as previously identified in the iterative process (Figures 3.6.A and 3.6.B).

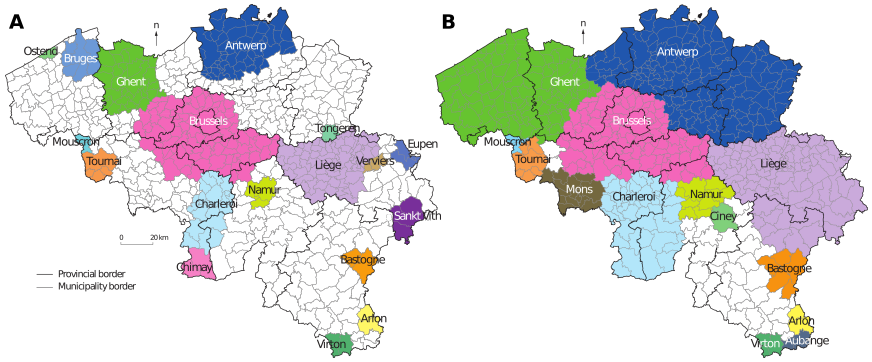


Figure 3.6. A) Residential basins based on the Anabel method – authorized inter cluster aggregation after 290 iterations - B) Residential basins based on the Anabel method – authorized inter cluster aggregation after 536 iterations.

Figure 3.5 thus shows a more advanced level of aggregation highlighting the principle areas of migratory influence in Belgium centred on polarities. The northern part of the country is divided into two large basins: Antwerp-Limburg and Ghent-West-Vlaanderen. The Brussels area emerges in the centre of the country, including the BCR, a part of the Flemish Brabant (Halle-Vilvoorde district), East Flanders province (Aalst, Ninove, etc.) and the entire Walloon Brabant province, the North of Namur province (Eghezée, Gembloux, etc.) and Hainaut (Ath, Soignies, etc.). This large basin extends beyond the linguistic border, fully permeable at this location. However, the impermeability is clearly demonstrated in the east between the provinces of Limburg and Liège. In Wallonia, the basin of Liège is a wide-area due to the lack of migration polarity in Luxemburg province[6]. The other basins expand around key Walloon polarities (Charleroi, Mons, Namur, Tournai) consistent with the urban hierarchy (Van Hecke *et al.*, 2009). The results obtained here present similarities with other hierarchical spatial frameworks, notably the urban regions specified by Luyten and Van Hecke (2007).

3.4.3 Louvain Method results

Communities are firstly detected within the migration exchange matrix, expressed in absolute terms. To obtain a partition of Belgium into a number of communities consistent with Anabel (8), we used the stability method (see Section 2.3.1 or, Delvenne *et al.*, 2013, for more details) and find that ρ must

[6]It is worth noting that a southern section of Luxembourg province is likely subject to the migratory influence of Luxembourg City, but these trans-border flows are not considered in this exercise.

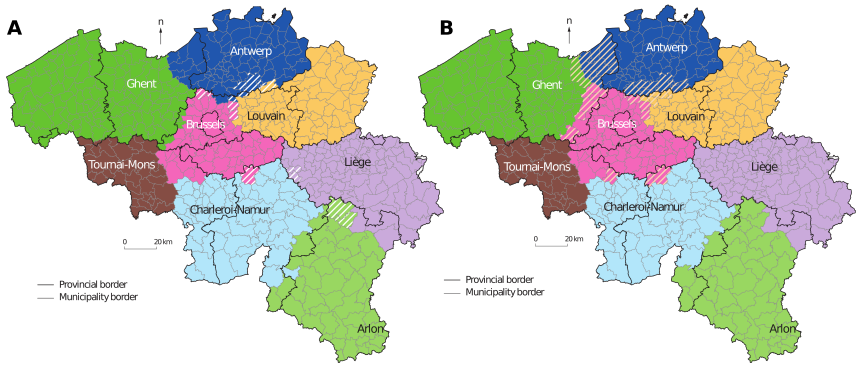


Figure 3.7. A) Communities detected using the Louvain Method (absolute date and $\rho = 1$) - B) Communities detected using the Louvain Method (relative data and $\rho = 1.25$).

be equal to 1.Results are mapped in Figure 3.7.A. The Louvain Method results in more homogeneous migratory communities (in size) than Anabel; these communities are structured around the main cities. Aside from municipalities around Brussels (urban sprawl) and the municipalities of Voeren and Comines-Warneton, the spatial organization of communities clearly shows the impact of the linguistic border: cross-border moves are rare. The municipalities that are hatched in Figure 3.7.A correspond to those which were not classified in the same community in over 10% of the cases (upon 1,000 consecutive applications of the method). These municipalities are primarily located on the edge of communities in zones of influence of several major cities (the zone between Antwerp, Brussels and Leuven).

Figure 3.7.B maps the results of the Louvain Method using relative flows with a value of $rho = 1.25$. Generally speaking, the results are not sensitive to the expression of the data (absolute or relative), apart from some municipalities at the edge of the two communities between Brussels and Ghent. Taking into consideration the absolute exchanges, the limit between the communities of Brussels and Ghent roughly matches the provincial border, while, in the study of relative linkages (Figure 3.7.B), the community around Ghent crosses the provincial border. Moreover, this zone is characterized by municipalities classified with a degree of uncertainty (hatched), proving the joint influence of the two cities (Verhetsel *et al.*, 2010). Accordingly, this lack of robustness helps to refine the limit between the communities of Brussels and Ghent. These hatched municipalities are therefore not linked exclusively to Brussels or Ghent, but are polarized by these two important urban centres (Van Hecke *et al.*, 2009). In general, the percentage of hatched municipalities is greater when the communities are detected with relative data (Figure 3.7.B) rather than absolute values (Figure 3.7.A). Substantial differences are found in the North

of the Brussels region and east of Antwerp. The Sint-Niklaas region is rather unstable in the partition based on relative exchanges.

3.4.4 Comparison of the results

Only the results from the Anabel method (relative linkages) and those of the Louvain Method applied to absolute migration values are compared here and illustrated in Figure 3.8 and Tables 3.2 and 3.3. Figure 3.8 illustrates the

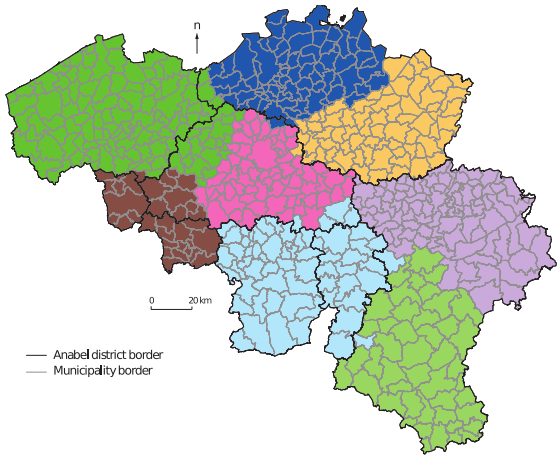


Figure 3.8. Comparison of partitions from the Louvain Method and Anabel.

differences between the delineations of the clusters obtained with the two methods. There is no exact match. As in the theoretical example, the Louvain Method seeks to generate similar size basins, whereas the Anabel method, applied without parameterized criterion, does not have the same objective. A key implication of this is a large basin for Antwerp and Liège with Anabel, whereas with the Louvain Method, Antwerp leaves room for a community covering Limburg and the east of Flemish Brabant, and Liège is separate from Luxembourg. Using the Louvain Method, Charleroi and Namur form a single community as well as Tournai and Mons. The Brussels community is also considerably smaller with the Louvain Method than with Anabel, because the latter approach takes into account the high weighted migration of Brussels since the start of the process, a migration weight which grows as the area gradually expands (Charlier *et al.*, 2016). In terms of similarities, the delineations match very well in the west and the north of the Liège basin, the north-west of the Luxembourg basin, the north of the Charleroi basin, between the basins of Charleroi and Mons and the basins of Antwerp-Limburg and Brussels. This resemblance between the two

chorochromatic maps is quantified by the normalized mutual information (NMI) value (see Equation 2.6). This index takes a value between 0 and 1, where 1 signifies that both partitions are identical (Strehl and Ghosh, 2002). The two partitions in Figure 3.8 have a normalized mutual information value of 0.77: the composition and delimitation of the basins and the communities are not identical, but there is a strong resemblance. As in the theoretical example, the Louvain Method seeks to generate relatively similar sized communities in terms of exchange flows, whereas the Anabel method, applied without parameterized criterion does not achieve that balance (Table 3.2). Table 3.3 compares results by means of an indicator (c) that measures the

Table 3.2. Composition of basins (Anabel) and communities (Louvain Method).

Anabel – 8 basins		Louvain – 8 communities ($\rho = 1.0$)	
Name of basin	**Number of municipalities**	**Name of communities**	**Number of communities**
Antwerp	150	Gent	126
Liège	134	Liège	88
Ghent	104	Antwerpen	80
Brussels	102	Leuven	69
Charleroi	40	Bruxelles	67
Namur	17	Charleroi - Namur	63
Mons	15	Arlon	44
Tournai	9	Tournai - Mons	34
Deviation	**51.13**	**Deviation**	**19.97**

importance of internal connections in groups (see Equation 2.2). For each partition (basin or community), the indicator measures the ratio between the weights of internal connections and the sum of weights for all the interconnections of that basin or of that community. The closer the value of c is to 1, the higher the ratio of internal connections. The values of c are generally quite similar in the results of Anabel (c average = 0.72) and of Louvain (c average = 0.75) with only two exceptions. In fact, the basins of Namur and Tournai have weaker c values in Anabel thereby demonstrating that the municipalities within these basins have a higher-than-average number of exchanges with other basins.

It should be reminded that the implementation of the two methods is performed using the inter-communal migration matrix, irrespective of intra-municipality migrations, since the Anabel method overlooks this aspect. Ignoring the intra-municipal moves (internal loops), means overlooking a large number of migrations, particularly in densely populated municipalities. If the Louvain Method applied with a value $\rho = 1$ irrespective of the intra-municipal migrations leads to a partition into eight basins

Table 3.3. Index *c* comparing basins and communities.

Anabel – 8 basins		Louvain – 8 communities ($\rho = 1.25$)	
Name of basin	**c**	**Name of communities**	**c**
Antwerp	0.88	Liège	0.86
Liège	0.87	Arlon	0.84
Brussels	0.84	Ghent	0.83
Ghent	0.82	Leuven	0.82
Charleroi	0.72	Charleroi - Namur	0.77
Mons	0.64	Antwerp	0.73
Tournai	0.54	Tournai - Mons	0.70
Namur	0.46	Brussels	0.69
Mean	**0.72**	**Mean**	**0.75**
Mean deviation	**0.13**	**Mean deviation**	**0.06**

(Figure 3.7.A), we are aware that the results and their interpretation are completely different when the intra-municipal migrations are considered: the optimal number of basins for $\rho = 1$ passes to 71 (see Thomas *et al.*, 2017) for the analysis and interpretation).

3.5 Conclusion

Two methods of spatial partitioning are applied on the same relational database: Anabel and the Louvain Method. The aim is to compare their advantages and limitations, while offering a critical geographical evaluation of the results based on a theoretical data set and a real world data set. The exercise therefore recalls the importance of method proficiency for interpreting results and a fortiori for decision making purposes in terms of governance.

Developed by INSEE, Anabel progressively aggregates places according to the maximal connection they foster and allows for step-by-step tracking. The reliance on maximal connection frequently identifies polarized structures and thereby a hierarchy of entities of different sizes. It is not so much the outcome of a final partitioning that is expected here, but rather the study of consecutive aggregations which makes it possible to understand the structure incurred by migratory relationships. However, the Louvain Method is heuristic and seeks to maximize the modularity of a network to define the different clusters of nodes among which migration exchanges are more frequent. Contrary to Anabel, this global method provides no information involving a possible hierarchy within migration exchanges. Accordingly, the obtained partitions are different (but clearly complementary), particularly in the multi-polarized territories where the population mobility is increasingly diffuse and complex.

In practice, the choice of the tool will largely depend on the objective: Anabel allows for an exploration of the mechanisms that constitute the basins, where the understanding of the spatial structure is equally as important as the method results. It is more suited to the analytical study of polarized territories, such as metropolitan areas and their extended zone of influence, and permits knowledge zoning (Théré and Séguy, 1998). It may therefore be used for the creation of supra-communal administrative entities structured around urban polarities (such as urban communities). The Louvain Method, however, gives an optimal spatial partition of relatively similar sized communities (in terms of exchange flows) while providing, for each node, a degree of classification uncertainty to the community it is assigned to. It serves to highlight spatial areas in which locations (nodes) have privileged exchanges, irrespective of previously defined polarities, without implementing a location in the method. It is therefore more suited to the implementation of a regionalization type territorial partition.

The danger of spatial partitioning without a good knowledge of the inputs and the method has here been demonstrated by means of two methods applied on an origin-destination data set. The choice depends upon the final objective of the partitioning; political control of egalitarian spatial partitions (Louvain) or analytical hierarchy (Anabel). Surely, it is the role of the scientist to be familiar with the field of possibilities in order to warn or guide policymakers according to the goals pursued.

III

Movements of people

4

Exploring job and residential basins in and around Brussels

If Chapters 2 and 3 explained the usefulness of the Louvain Method, the objective of this Chapter is to show how the method enables to refresh the geography of movements of people (migration and commuting movements obtained from the Census) first at the country level (Section 4.3.1) and further in and around the Metropolitan area of Brussels (Section 4.3.2).

While neither distance nor the characteristics of the places are inputs of the method, a mosaic of "interaction fields" is delineated, made out of contiguous places whatever the scale of analysis. Commuting between residence and work, and changing place of residence do not lead to the same spatial basins reminding that great caution should be put on the choice of the data and on the hypothesis to be tested. The nature of the movement and hence the processes matter.

This Chapter is based on works already published in *Brussels studies* and *Belgeo* (see Adam *et al.*, 2017; Thomas *et al.*, 2017)

4.1 Introduction

From an administrative and political standpoint, Brussels is often referred to as the Brussels-Capital Region (BCR), yet the city spreads out over a much larger area whose boundaries are not officially and scientifically fixed (Dujardin *et al.*, 2007; Thomas *et al.*, 2012; Boussauw *et al.*, 2013; Vandermotten, 2016). Many different criteria and methods are used to define the Brussels metropolitan area, based on one or several socioeconomic and/or morphological variables, and on methods which are essentially related to the value thresholds of the variables used (Dujardin *et al.*, 2007; Thomas *et al.*, 2012). Economists and geographers agree that cities are characterised by high density in terms of population, buildings and urban functions with respect to their environment. Consequently, the agents (firms and households) are close and interact (Anas *et al.*, 1998; Fujita and Thisse, 2013). Cities have a concentration of a wide variety of agents, and these agents interact within economic and social networks. Proximity plays an important role in the formation of socioeconomic and human ties between agents (Ioannides, 2012) and thus leads to "interaction fields" (Fujita and Thisse, 2013). Delimiting urban areas and defining their composition is not a new topic in urban analysis, but there is no consensus regarding the methods used and the results obtained (Duranton, 2015; Dujardin *et al.*, 2007).

Nowadays, the way of conducting the Census in Belgium has totally changed and the data needed for repeating the abovementioned urban delineation are not easily available anymore. In recent years, newly collected (big) data allow us to measure and reveal new aspects of the urban complexity and especially the pulses of the territory in time and space and the interrelationships between people and places. The development of ICT technologies, of increasing data storage and computer capacities indeed enable to collect very large volumes of data and hence better follow people and firms in their everyday activities. Hence, we know where people are (GPS in smartphones) and with whom they are in contact (social media, phone data, etc.). The ICT data constitute a new opportunity for charactering nodes (X_i), and the interrelationships between them (W_{ij}). In this sense, the partition of the Belgian telephone territory has become a well-known example based on the frequency of communication between municipalities (Blondel *et al.*, 2010), undoubtedly reflecting a geography of phone basins, closely related to the daily urban system. Comparing the phone map with the "city regions" reveals challenging similarities, but also differences and complementarities (Figure 4.1). If big-data has the advantage of opening new research opportunities, they are often unfortunately not always in the public domain (privacy rules) and need a very clear understanding of what they are really measuring Kitchin (2013). Big-data and communication technologies have also led to the development of new methods in Data Sciences for

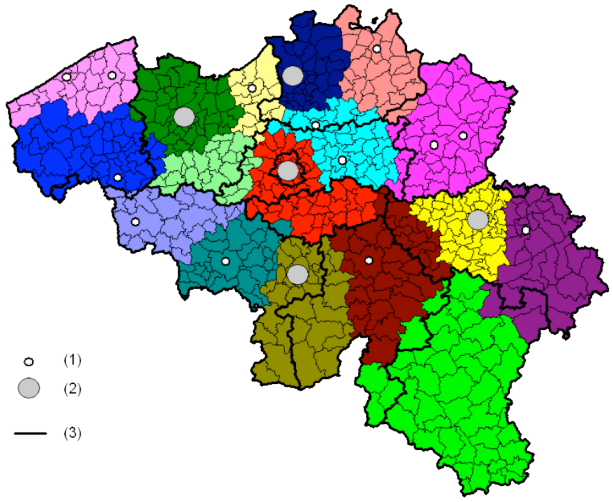


Figure 4.1. Phone basins based on cell phones billing addresses (Blondel *et al.*, 2010) (1) regional city, (2) major city , (3) provincial border.

optimally detecting groups of individuals (people, firms, etc.) that interact (see e.g. Blondel *et al.*, 2008); they have the advantage of letting the data speak for themselves (data driven models). While these individuals are geolocated, the spatial projection of the community is related to the "interaction fields". Despite their limits, these new methods and data have turned out to be effective tools for analysis, visualisation and urban planning (Sagl *et al.*, 2014; Hao *et al.*, 2015). These new methods hence offer interesting opportunities to study interactions fields in and around cities and hence better understand the economics of agglomeration (see e.g. Fujita and Thisse, 2013) and verify theory driven models. It is now time for geographers to catch up this new opportunity and link these new results to spatial knowledge and theory critically.

Before using this kind of big-data (see Chapters 5, 7 and 6), this Chapter constructs a new geography of Belgium and Brussels using very traditional relational data provided by the last census (changes of residence, commutes to and from the workplace). It aims to show which places are closely linked in and around Brussels, thanks to the delineation of the "interaction fields". This Chapter is hence an extension of the analyses already proposed in Chapter 2 and 3 with a deeper understanding of the geographical outcomes instead of methodological explanations. It suggests areas based on relational data (W_{ij}), and provides a complement to the traditional territorial dynamics

often based on the similarities between the characteristics X_i of places (Van Hamme *et al.*, 2016). What are the strengths and weaknesses of these interaction approaches for urban analysis? Are the areas made up of adjacent places? Are the interactions between people or places in and around Brussels influenced by political (regional, linguistic) or physical boundaries (such as the canal or the Sonian Forest), and/or by other types of barrier (Dobruszkes and Vandermotten, 2006)? Does the shape of these areas depend greatly on the data used?

4.2 Data and method

4.2.1 Data: migration and commuting

We here limit ourselves to census data, and leave the debate about ICT/big-data aside: ICT data have high potentials, but also strong limitations especially when dealing with geolocations. This question will be raised in Chapter 5, 7 and 6.

Very typically, we have first used home-workplace travel (commuting) and residential migrations (moves) recorded during the last census (2011 - Census11), with all of the advantages and disadvantages of a census (Thomas *et al.*, 2009). These OD matrices have been available every ten years since 1981, and have the advantage of corresponding to definitions (very well documented, see e.g. the Statbel website) which are clear and stable over space and time, even though, since 2011, their construction is no longer based on the census survey but based essentially on existing data collected by different Belgian and Regional administrations. The transition from survey to administrative census data questions the temporal comparability for many variables and raises major criticisms in the data and their definition.: levels of uncertainty related to the definitions of the 1) workplace - head office used instead of the actual workplace, multiple or non-fixed workplaces, certain categories of workers not taken into account such as international civil servants (Verhetsel *et al.*, 2009) - or 2) of the place of residence - declared and not actual, multiple moves in a year (Eggerickx, 2013).

We therefore expect to define groups of places forming employment basins (commutes to and from the workplace) and migratory basins (moves). Commuting data rely on the coupling the National Register and the Social Security Register (BCE) and migration data are based on the National Register that centralises information collected by the municipalities. Commuting or migrating with another country is unfortunately too inaccurately spatially coded to be considered here; our analyses are hence limited to movements within Belgium although foreign countries could have an impact on the partitions delineated. Both migration and commuting

movements are officially available at the scale of the municipalities. Commuting is also available at the level of the statistical sectors (neighbourhoods); these latter are also used here to show how far the level of aggregation influences the final results. Hence, we end up with three data sets: migrations and commuting between the 589 municipalities, and commuting between the 19,782 statistical sectors.

4.2.2 Method

To detect communities of places that are tightly connected, the Louvain Method (see Chapter 2 for further details) is applied on both datasets and is conducted on the scale of Belgium. Geographical discussions are proposed at the Belgium scale in Section 4.3.1, while the Section 4.3.2 proposes to make a zoom on the area of Brussels. Results are illustrated in the form of maps with an arbitrary choice of colours. The thin lines correspond to the municipalities while the thick lines represent the provincial boundaries. The very small meshes which do not include enough observations have been left blank. We know that the Louvain Method is sensitive to the order of interpretation of the nodes; in order to avoid this bias, it is applied 1,000 consecutive times while randomly modifying the order of the nodes between each application, which then allows the strength of the results to be measured: the percentage of times which a node belongs to the same community reflects the tness of the result and the independence of the order of places. This percentage is illustrated by choropleth maps (statistical maps representing relative quantities associated with geographical areas by means of a scale of graduated shades) and/or by hatching on chorochromatic maps (coloured maps representing a nominal variable, in this case belonging to a community). On certain maps, we have also indicated the nodes where there is the highest concentration of connections and which constitute centralities (black dots on the map). To do this, we have taken the number of connections for each Belgian node and have applied the Jenks method to the distribution in order to pinpoint the places with the most connections. Each partition is accompanied by simple statistics such as the number of places in each community and the percentage of places whose classification in the 1,000 applications is unstable. Finally, the similarities between maps are quantified by NMI (Normalized Mutual Information) which measures the quantity of information contained in a map which is also contained in another map (see Equation 2.6 presented in Chapter 2). This index has the value 0 when two maps are totally different, and 1 when they are identical (Ana and Jain, 2003).

4.2.3 Defining the extension of Brussels

It is often difficult to define the boundaries of an area under study and its partitioning, particularly as they relate to a city for which there is no consensus as to its exact boundaries and which spreads out over three Regions (Brussels-Capital, Flemish Region and Walloon Region). In order to avoid any disputes, this Chapter proposes to use the former province of Brabant as the area under study. This entity was created in 1815 and today is divided into three entities: the Brussels-Capital Region (1.2 million inhabitants), the province of Flemish Brabant in the north (1.1 million inhabitants) and that of Walloon Brabant in the south (0.4 million inhabitants). The province of Brabant includes most of the Brussels urban area (Thomas *et al.*, 2012) as well as Leuven, whose position in the metropolitan area is debated. While the maps proposed here constitute a close-up look at Brabant, the analyses were carried out for all of Belgium for migrations and commuting, which allows side effects to be avoided in a country with a dense urban network, without modifying the general appearance of the partitioning

The proposed illustrations were made mainly at the scale of municipalities in order to be able to compare the results. This choice is justified by the fact that a municipality is the most specific geolocation obtained for moves and is therefore the smallest common denominator of the analyses. The smallest spatial units for commuting are the statistical sectors (numbering 3,395).

4.3 Results: spatial partitions

4.3.1 The case of Belgium

Job basins

The problem is first to find the optimal ρ value, if any. As already developed in Section 2.3.1, we use the method of Delvenne *et al.* (2013) and the Figure 4.2 illustrates this search. As already mentioned, $1,000$ runs are applied for each value of ρ. The similarity between each set of $1,000$ runs (partitions) is computed by means of a Variation of Information (noted VI) and iteratively computed for each value of ρ. The optimal partition corresponds to the minimal value of VI. Using $\rho = 1$ corresponds to the standard version of the Louvain Method; it is a default value for the method, corresponding to a situation where the optimization of the modularity is not modified. When ρ is smaller than 0.02, VI values are always equal to 0 because the Louvain Method cannot detect communities in the network; in this case, the number of communities is equal to the number of nodes (589). When ρ is very large (larger than 5), the Louvain Method has merged all the nodes in one single community. It is hence impossible to observe a positive

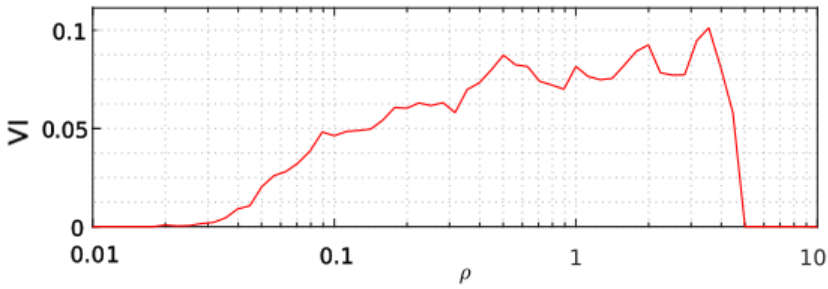


Figure 4.2. Variation of Information (VI) versus ρ values for commuting movements between municipalities.

value of VI. Figure 4.2 clearly shows that there are 2 main and clear local minimal values of VI: $\rho = 0.3$ and $\rho = 0.9$. We also know that the best partition has more (weighted) edges within the communities - internal edges - than between communities - external edges - (Newman, 2004). Figure 4.3 shows the proportion of internal and external edges, for increasing values of ρ and hence for decreasing values of the number of communities. The proportion of internal edges is higher when $\rho = 0.9$ than when $\rho = 0.3$. Based on the VI method (also called Stability Method) and on the proportion of internal edges, we hence consider that the best community structure (optimal) is found when $\rho = 0.9$.

Once optimal values are correctly determined, Belgium is optimally partitioned in about 10 job basins which is much less than the 47 defined by de Wasseige *et al.* (2000) but more than the five large cities often defined as job centres (see e.g. Van Hamme *et al.*, 2011, and Figure 1.1). Contrarily to other publications using thresholds (Marissal *et al.*, 2006) or other methodologies (Charlier *et al.*, 2016; Adam *et al.*, 2018a), the method allocates all municipalities to a community. Results of the Louvain partition are mapped at the level of municipalities (Figure 4.4.A) and at the level of the statistical sectors (Figure 4.4.B). Comparison of the two maps is straightforward: almost the same number of communities (respectively 10 when nodes are municipalities, and 11 for statistical sectors) is observed and the two spatial partitions clearly look alike. Data aggregation little impacts the partitioning of Belgium in terms of commuting movements. Note that uncolored statistical sectors (Figure 4.4.B) are due to the absence (or too small numbers).

The limits of the communities correspond quite well to provincial borders: commuters tend to stay in the same province at some exceptions mainly located in the vicinity of the two largest cities (Brussels and Antwerp).

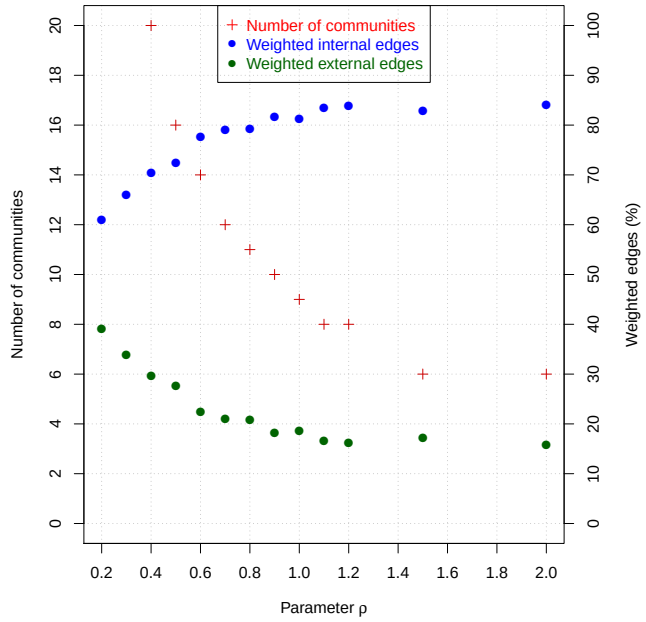


Figure 4.3. Proportion of internal and external edges and number of communities for different values of ρ for the 2011 commuting movements (at municipality level).

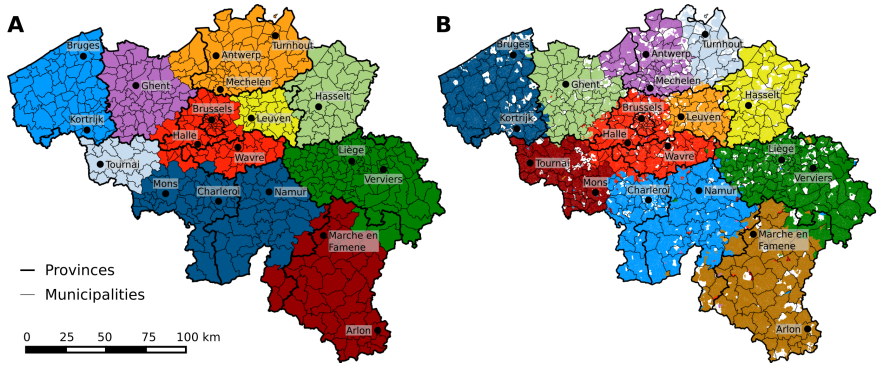


Figure 4.4. Optimal partition in communities when nodes are municipalities (A – $\rho = 0.9$) and statistical sectors (B – $\rho = 1.0$) - 2011 commuting data.

Communities around these two big cities are indeed large, and they step over regional/provincial borders: Brussels extends to Halle-Vilvoorde and encompasses the entire Brabant wallon stepping over the linguistic border, but it does not include the area of Leuven which is a job basin on its own. The community centred on Antwerp includes Waasland (across the provincial border) while the Eastern part of the province of Antwerp is a job basin on its own (Antwerp Kempen) when we work at the scale of the statistical sectors. Also note that in Wallonia, the provinces of Liège and Luxembourg correspond to clearly defined communities, while the limits of the provinces of Hainaut and Namur do not correspond to that of the communities Charleroi and Namur form one job basin, while Mons and Tournai form another one).

These spatial observations are confirmed by a simple visualisation of the raw commuting flow. Figure 4.5 presents the number of movements made between pair of municipalities where these municipalities are organized on ascending order on both X and Y axes. Strongest exchanges are hence found on and around the diagonal of the matrix, expresses that strongest movements are made within the "neighbourhood". Squares are quite clear and correspond to provincial limit or to two historical arrondissement limit (Leuven and Halle-Vilvoorde). It is also surprising to see that the municipalities belonging to the RBC have stronger exchanges with the arrondissement of Halle-Vilvoorde than with the one of Leuven. It seems that former (or less former) historical organizations have still a strong effect on the commuting movements (the former arrondissement of Brussels-Halle-Vilvoorde). This kind of visualization is also quite helpful to better understand how the various flows are organized and embedded within space and how the communities detected at different ρ values have a strong hierarchy (see Chapter 2 and Figure 2.6 for several job basins obtain with multiple ρ values). Let us remind here that location is not an input of the method; the intensity of the commuting fluxes is the only input. Hence, commuting interactions in Belgium lead to a very clear structure dominated by the contiguity of the basins (small distances are more important than large ones), and an imprint of the provinces; moreover, the urban hierarchy (Brussels and Antwerp) and transportation infrastructure (Mons-Tournai, Namur-Charleroi) also intervene.

The advantage of running $1,000$ times the Louvain Method for each value of ρ is that each node can hence be characterized by the percentage of runs for which it is associated to the same community, hence measuring the robustness of the partition. Choropleth maps of the robustness of node in the partition can hence be drawn (Figure 4.6). Note that it is hard to get the same discretisation of the values observed at the municipality level (Figure 4.6.A) and for statistical sectors (Figure 4.6.B) because of the large differences in the number of nodes

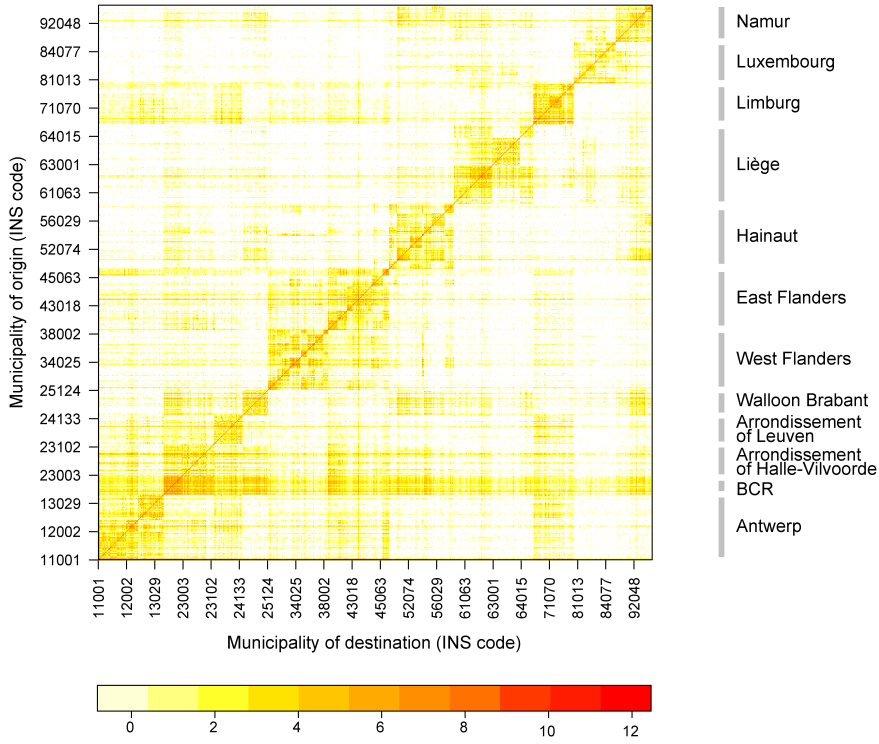


Figure 4.5. Square matrix of commuting movements appearing in Belgium.

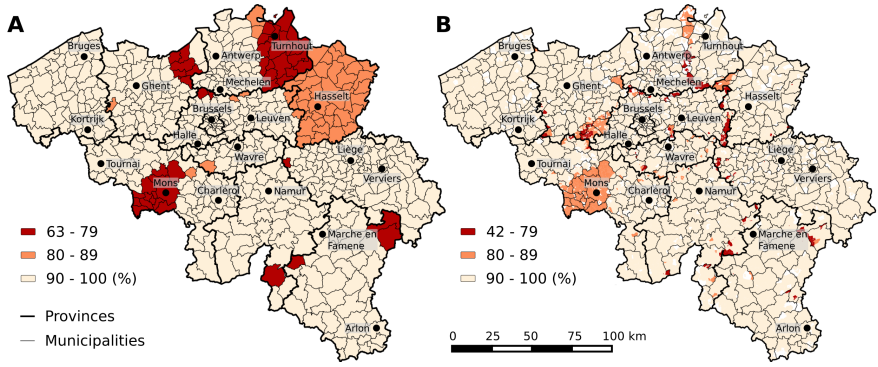


Figure 4.6. Robustness of the partitions reported in Figure 4.4: % of times a node belongs to the same community in the 1,000 runs - 2011 commuting data.

and in the statistical distribution of the variable. By visually comparing the two maps in Figure 4.6 we see that the coloured surface is much smaller and lighter for statistical sectors and it mainly corresponds to statistical sectors located at the limits of the communities.

Furthermore, the role of ρ in the final partitioning of Belgium is illustrated by 4 maps in Figure 4.7. When $\rho = 1$ (default option), 9 communities are delineated. The job basin of Brussels sprawls out of the BCR into the arrondissement of Halle-Vilvoorde and encompasses Brabant wallon and even extends to Flanders (West) and Hainaut (South-West). The borders of the communities of Luxembourg, Liège, Antwerp and the two Flanders correspond at some exceptions (such as Waasland where recently major expansions of the Antwerp port are located) to the provincial limits, while Leuven and Limburg merge and so do also the job basins of Charleroi and Namur. For these two latter cases, rail/road infrastructure might strongly explain the merge. When ρ is smaller (0,7), 11 communities are delineated. Increasing ρ of 0,1 unit ($\rho = 0,8$) does not change the number of communities (11) but the delineation is slightly different (Hainaut shrinks, Namur and Brussels enlarge). In both cases ($\rho = 0,7$ and 0,8) there are more communities in the Northern part of the country than in Wallonia in accordance with the structure of the urban network (more job centres in the North, closer to each other; for historical and geographical reasons, denser city network in the North). When choosing a much larger value of such as $\rho = 2$, 6 communities appear: Brussels, West and Oost Vlaanderen (together), Antwerp (including Waasland), Limburg, Liege and Hainaut-Namur. Hatched municipalities correspond to less robust allocations: the classification of these nodes is less certain because they share links with several attracting Belgian job centres (let us here remind that no foreign job centre was considered – see

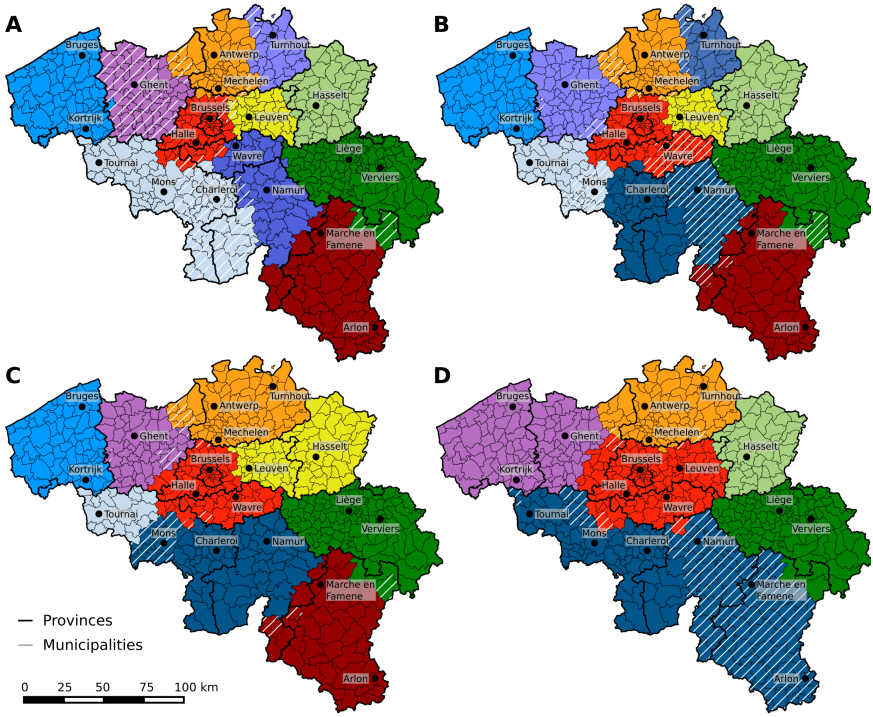


Figure 4.7. Commuting communities for different values of ρ (A = 0,7 - B = 0,8 - C = 1,0 - D = 2,0). Hatched municipalities are those where the robustness is $< 90\%$.

Section 4.2.1).

As concluding comments of this Section, we insist on 3 points about commuting basins: (1) as expected, the larger the value of ρ, the smaller the number of communities. Hence, the search of an optimal value is clearly required in order to avoid misinterpretation. (2) All communities are made up of contiguous municipalities meaning that links between contiguous municipalities are always stronger than between distant places: space hence matters. Gravity type relationships extraordinarily still matter despite transportation and ICT infrastructures: workers live and work in the same area. (3) The limits of the communities fit quite well with those of the provinces as well as that of the linguistic border, at the exception of the southern part of the community of Brussels that sprawls into the province of Walloon Brabant and steps over the linguistic border.

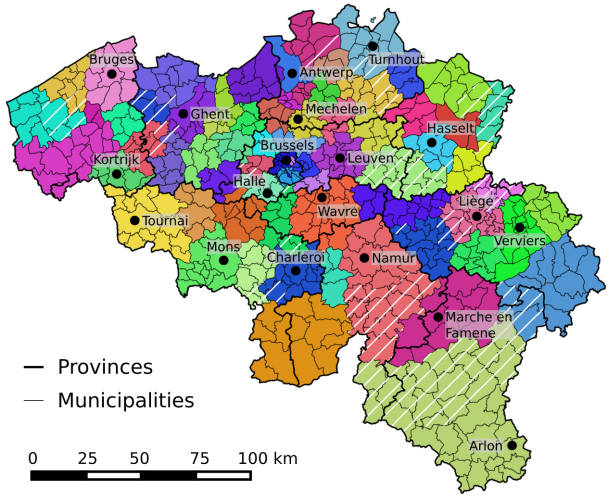


Figure 4.8. Migration communities when $\rho = 1.0$. Hatched municipalities are those where the robustness is $< 90\%$.

Residential basins

Let us now take another O-D matrix and apply the Louvain Method on the number of migrations between Belgian municipalities. When considering the standard Louvain Method ($\rho = 1$), 71 migrations fields are delineated (Figure 4.8): this means that most people decide to change their place of residence for another place in the close vicinity and that migration basins are very small compared to commuting basins. Commuting and migration correspond to two different spatial choices processes. Another difference with commuting maps is that the linguistic (regional) border is much more impermeable. This does not mean that nobody moves from Wallonia to Flanders or inversely, but those moves are anecdotic compared to other fluxes.

Based on the Stability Method, the optimal value of ρ is this time equal to 3, which corresponds to 24 migration fields; this is still much more than the optimal 11 commuting basins. Finally, we force the Louvain Method to run with a ρ value equal to 8 in order to get 11 migration fields (Figure 4.9). We end up with a partition resembling that with $\rho = 1$ computed for commuting at the exceptions of Waasland and Leuven that form specific migration fields and also some municipalities at the border of the delineations that are differently allocated. Hence, the Louvain Method run on migration data with$\rho = 3$ corresponds to an optimal partition, but with a relatively large percentage of robustness (hatched surfaces), while the solution with $\rho = 8$ is not considered as optimal but spatially more robust (less hatched surfaces). This robustness

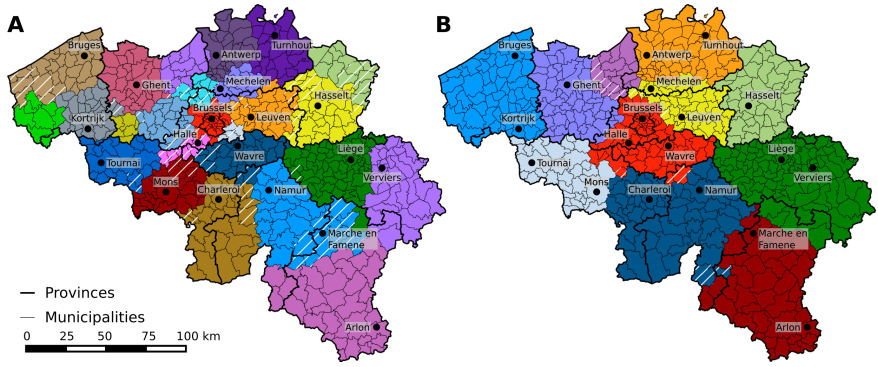


Figure 4.9. Migration communities when $\rho = 3.0$ (A) and $\rho = 8.0$ (B). Hatched municipalities are those where the robustness is $< 90\%$.

has to be related with the number of communities detected: there are more communities when $\rho = 3$ than when $\rho = 8$.

Comparison

Differences between migration and commuting fields are here quantified as well as their resemblance with provincial delineations. Table 4.1 refers to the pairwise comparisons of the chorochromatic maps by means of the NMI. Values taken by the NMI index confirm the visual observations: they are high between commuting and migration fields, the most important differences being observed for ρ values equal to 1.0 (NMI=0.640). The maximum value of the NMI is reached for commuting and migrations fields observed with a ρ value of 0.8 for commuting and 8.0 for migration: in this case 88.8% of the partitioning is similar (NMI=0.888). The last column of Table 4.1 gives the NMIs between the provincial delineation and the computed migration/commuting fields. When the number of communities is comparable to the number of provinces (11), NMIs are generally larger than 80% meaning that maps largely look alike, confirming our visual analysis.

Comparison is pursued in Table 4.2. We see that for migration data, the number of communities is much larger (diversity) and hence the size of the communities is smaller. In other words, and geographically speaking, distances are smaller when selecting a new place of residence: when households move they want to keep their relatives, their social relationships, neighbours, sports or collective activities and hence do not move far away. When selecting a job place (commuting), workers are ready to travel daily a longer distance; note that this trend is supported in Belgium by several policy tools such a reimbursement of commuting travel, company cars, etc.

Note also that the second line of Table 4.2 refers to the robustness of the

Table 4.1. NMI values between pairs of partitions (municipality level).

	Com. ($\rho = 1.0$)	**Mig.** ($\rho = 1.0$)	**Mig.*** ($\rho = 3.0$)	**Mig.** ($\rho = 8.0$)	**provinces**
Commuting* ($\rho = 0.9$)	0.896	0.688	0.766	0.872	0.861
Commuting ($\rho = 1.0$)	1	0.640	0.711	0.819	0.848
Migration ($\rho = 1.0$)		1	0.823	0.715	0.660
Migration* ($\rho = 3.0$)			1	0.792	0.731
Migration ($\rho = 8.0$)				1	0.834

* optimal solution in Figure 4.3, Com = commuting, Mig = migration

solution on the 1,000 runs is another type of information, different from modularity. For instance, migration data analysed with a $\rho = 3.0$ corresponds to an optimal solution, but with a relatively large percentage of robustness, while solution with $\rho = 8.0$ is not optimal but more robust. This once again illustrates that one map (one partition) should always be interpreted carefully, taking into consideration the methodological choices made the researcher. Mathematical optimality and robustness do not always converge. Data may indeed speak by themselves, but what they reveal is significantly framed by the researcher himself (Miller, 2017).

Table 4.2. Comparing solutions for partitions based on municipalities.

	Com* ($\rho = 0.9$)	**com** ($\rho = 1.0$)	**Mig** ($\rho = 1.0$)	**Mig*** ($\rho = 3.0$)	**Mig** ($\rho = 8.0$)
# communities	10	9	71	28	11
% of hatched	17.00	5.85	16.87	15.15	2.58

* optimal solution in Figure 4.3, Com = commuting, Mig = migration

Let us go back to Figure 1.1 that gives the delineation urban agglomerations and the importance of the city, and so implying that not all municipalities are allocated to a centre (white on the map). The objective is hence different from that developed in this chapter. Hence, all municipalities are allocated to a community and no centrality is a priori defined. Hatched surfaces reveal places for which the allocation to a community is less certain and it is a posteriori possible to define centres (Beckers *et al.*, 2018; Adam *et al.*, 2017). The objective of the map reported in Figure 1.1 is hence different from that

pursued in this chapter, but both bring an interesting and complementary viewpoint on the geography of Belgium.

4.3.2 The case of Brussels

Job basins

With employment basins come employment centres. The Louvain Method does not require a specific central location: it classifies the individuals in a community based on the intensity of the connections between them, independently of the location. In other words, employment centres are not designated a priori, and the notion of employment basins could become debatable. However, the calculation of centralities a posteriori (see Section 4.2.2) highlights most of the municipalities of the Brussels-Capital Region and the adjacent municipalities such as Dilbeek, Asse, Grimbergen, Vilvoorde, Machelen and Zaventem, as well as municipalities such as Halle and Leuven and the employment centres in Walloon Brabant (Wavre, Ottignes-Louvain-la-Neuve, Braine-l'Alleud and Nivelles).

Contrary to expectations, the level of spatial aggregation of data affects neither the optimal number nor the global spatial structure of the communities based on daily homeworkplace travel (Figures 4.10.A and .B). The communities calculated based on statistical sectors simply have more detailed outlines, and many sectors are not classified in the absence of observations. Regardless of the level of aggregation, two areas emerge, made

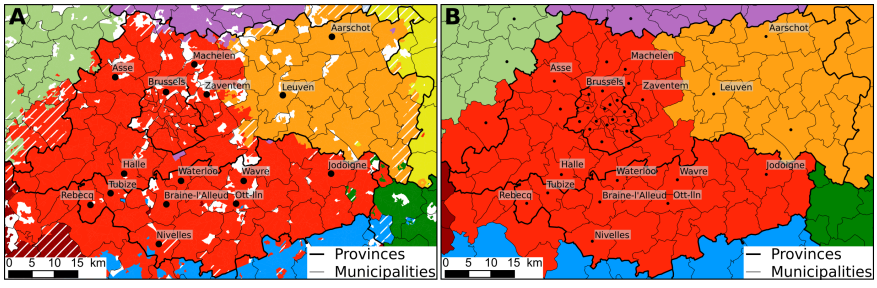


Figure 4.10. Communities based on commutes to and from the workplace in 2011, on the scale of the statistical sectors for $\rho = 1$ (A) and municipalities for $\rho = 1$ (B).

up of places which are adjacent to one another: on the one hand, the employment area of Brussels which includes the Brussels-Capital Region, the district of Halle-Vilvoorde and Walloon Brabant, and on the other hand, the area of Leuven which covers the district with the same name, with the exception of some bordering municipalities such as Tervuren, Huldenberg and

Kortenberg, which are associated with the Brussels area. Let us underline the fact that Leuven forms a community in itself, not because there is no relationship between this area and the one which is centred on Brussels, but because the places which form the Leuven Community exchange more commuters between themselves than with the rest of the Belgian municipalities (this therefore does not prevent exchanges between the red and orange municipalities on the maps in (Figures 4.10.A and .B).

Figure 4.11 illustrates graphically the exchanges between the communities. In order to be legible, the exchanges between communities 1 (Leuven) and 2 (Brussels) are illustrated, while the exchanges between the other communities in Belgium are included in a virtual group named *a*. This circular graph clearly highlights the way in which the Louvain Method classifies the municipalities into communities: communities have a high level of internal exchanges and a low level of external exchanges.

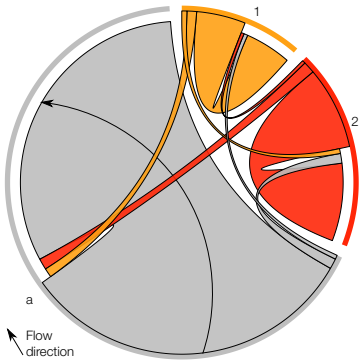


Figure 4.11. Flow of commutes between communities 1 (Leuven), 2 (Brussels) and all of the other Belgian communities (a).

As a secondary visualization, Figure 4.12 maps the main commuting flows appearing within the former province of Brabant. Two areas are clearly delineated, one centred on Brussels and the other one is dominated by Leuven. Exchanges between these two groups are not intense. Not surprisingly, the area dominated by Brussels is made by several small centres (Halle, Wavre, Nivelle). Similarly to Figure 4.5, the matrix presented in Figure 4.13 shows the movements measured between pairs of municipalities. Squares that are observed on the diagonal correspond to pair of municipalities that are exchanging a lot of commuters. It is clear that the RBC attracts most of the workers at the exception of those living in the arrondissement of Leuven.

Although the municipality of Nivelles is included in the community of Brabant, it demonstrates a certain lack of robustness in terms of classification, which results in a tendency towards exchanges shared with the

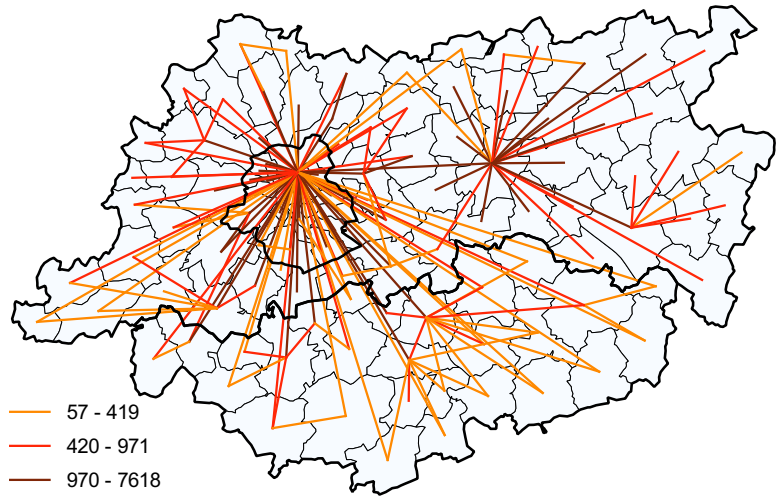


Figure 4.12. Principal commuting movements appearing in the former province of Brabant.

area of Hainaut (Figures 4.10). In keeping with the history of employment and the railway infrastructure (Verhetsel *et al.*, 2010), the area of Brussels is more extensive to the west and southwest, going well beyond the provincial and regional boundaries. Finally, let us note that the provincial boundaries correspond quite closely to the boundaries of the communities, confirming Thomas *et al.* (2017): although these boundaries are debated and tend to be removed, they appear clearly in the functioning of the labour market. In the same line, the linguistic boundary is not permeable in the eastern part of the area under study, whereas from Rixensart/Overijse to the extreme west of the area under study, the linguistic boundary does not at all constitute a barrier in terms of commutes to and from the workplace.

Let us bear in mind that each map is the result of 1,000 applications of the method to a data set and that it is therefore possible to associate a degree of certainty of the classification to each place. Figure 4.10 (as well as Table 3.1 below) shows that less than 5 % of the nodes are classified in an uncertain manner; they are located mainly at the edge of the province (Londerzeel and Kapellen-op-den-Bos in the north, and Nivelles in the south).

In conclusion of this Section, analysing commuting movements shows that the partitioning into communities leads to strong results (few areas with hatching), that only two basins emerge in an optimal way, that the mesh size (statistical sectors, municipalities) has little or no impact on the general

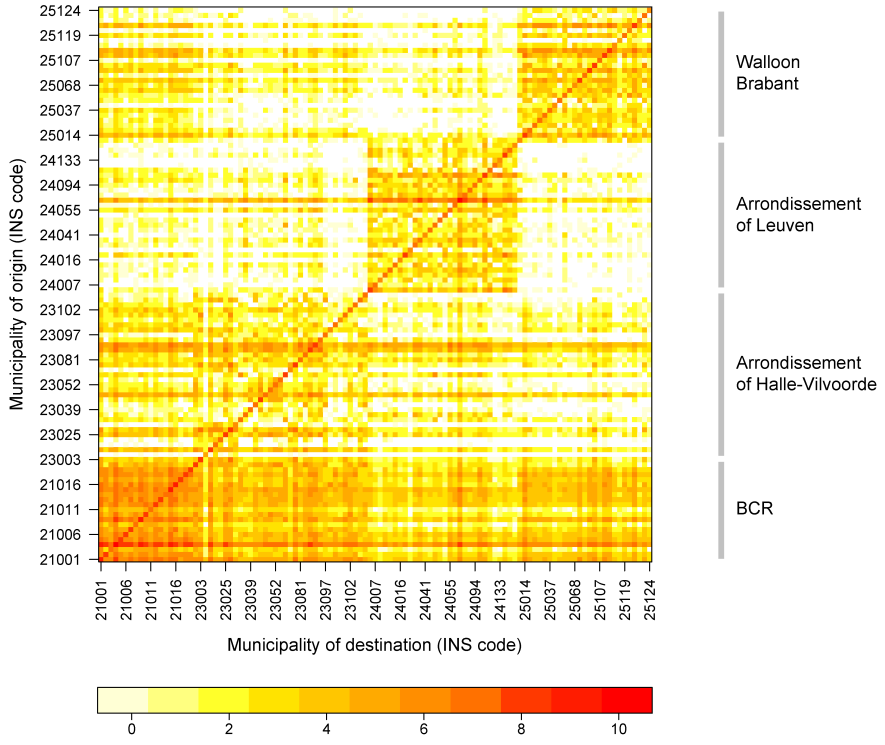


Figure 4.13. Square matrix of commuting movements appearing in the former province of Brabant.

aspect of the partition, and that the communities are made up of places which are adjacent (distance has a strong impact on the behaviour of commuters).

Residential basins

Residential migrations, commutes to and from the workplace: two types of movement of people between places which respond to totally different mechanisms of spatial choices and territorial practices. It is therefore not surprising that the partitions obtained are different (NMI migrations-commutes: 0.64). Results of the analysis are summarised cartographically in Figure 4.14. The breaking up of Brabant is more clearer (21 communities instead of 2 for commutes) and results in small groups of municipalities where a change of residence between them is very frequent. Another piece of information – and therefore a partition – is also a constant: the groups are formed of adjacent municipalities. The communities are of a very small size (from 2 to 16 municipalities): the moves in and around Brussels are of a very short distance. While the communities overlap the exterior envelope of the province of Brabant, only two communities cross the linguistic boundary: thus, Sint-Genesius-Rode/Rhode-Saint-Genèse are part of the same community as Braine-l'Alleud and Nivelles, of course with a slight amount of uncertainty (hatching), and Herne is part of the community centred on Soignies/Braine-le-Comte/Enghien. The Brussels-Capital Region is divided into four communities: (1) a northwest part, (2) a south/southeast part which also includes Linkebeek and Drogenbos, (3) a small community made up of Saint-Josse-ten-Noode, Evere and Schaerbeek, and finally, (4) a community which includes the two Woluwe and continues towards Tervuren and Zaventem (Van Criekingen, 2006; Vandermotten *et al.*, 2009a; De Maesschalck *et al.*, 2015).

In terms of moves, distance, territory and language have a stronger impact on spatial choices than commutes. The nodes whose classification is uncertain represent 17 % of the total, i.e. a much higher percentage than in the case of commutes (5 %). This therefore means that the partition is weaker than that of commutes (hatching) and that the places are divided between several polarities. The geography of moves can therefore not be used instead of commutes: it involves a completely different spatiality (people may choose a job far from home, from among several employment centres, but when they move they might try to stay near friends, schools, leisure activities, etc.).

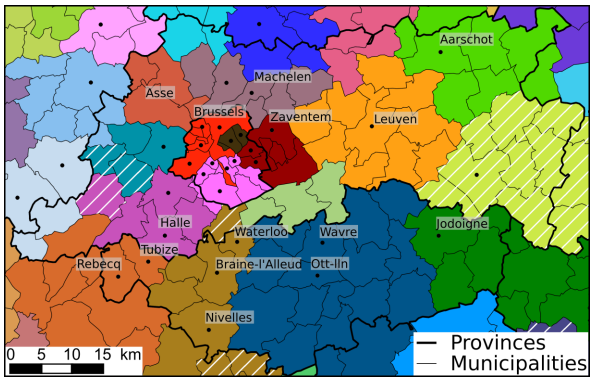


Figure 4.14. Optimal division into communities ($\rho = 1$), based on residential migration data from 2011.

4.4 Conclusion

The detection of communities in (large) relational databases is a recent approach whose applications in a spatial context are still rare and in their infancy (Zhong *et al.*, 2014). A "new" geography of Belgium and of the Metropolitan area of Brussels is here proposed, relying on two OD matrices (commuting and migrations) and a network based method (the Louvain Method) that optimally clusters places without predefining central places or other characteristics of the locations. It here proved to be quite useful and informative, especially when adapting the objective function with recent methodological developments (Delvenne *et al.*, 2013). With such methodological choices, one can easily state that *data speak by themselves*: we indeed obtain partitions without fixing any a priori thresholds, not even including geocoding in the method. We however showed that behind this idealistic methodological viewpoint, there are still some choices to be made by the researcher and hence there is no unique and indisputable solution for defining optimality (choice of ρ, robustness of the allocation of a node to a community). Optimality and robustness are two different but mixed concepts that should be weighted when providing partitions to the decision maker, and results have always to be interpreted with a geographical knowledge of the study area confirming a paper of Miller (2017). The cartographies proposed show communities of places with strong relationships (W_{ij}) and allow the space in Brussels to be defined and partitioned. They illustrate the complex urban spatial reality (in Brussels), and call to mind the importance of having a proper understanding of the nature of the variable used before drawing operational conclusions.

Both datasets provide a different mosaic of basins reminding how important it is to control data definition in any quantitative analysis. Due to their definition, commuting and migration lead to different spatial structures (number, composition of communities). The shape and the number of areas depend greatly on the data used, and it is important to question the nature and limitations of data and methods. However, the maps refer to relatively strong and often known spatial structures, as though the space and in particular the urban network had reached a certain *equilibrium*, suggested by (Fujita and Thisse, 2013) and discussed at length by (Pumain, 2017).

Commuting data reveal a few large interaction fields confirming that commuters travel daily a long distance starting from their place of residence (communities are always made of contiguous municipalities). This may be explained by 1) politics in Belgium that allow and encourage the reimbursement of home-work travel expenses reinforce people to search for a job within a larger range of distances, and/or 2) economic and residential areas do not match, pushing workers to travel more to reach their work places. Results obtained at the level of the statistical sectors are exactly the same as those obtained at the level of municipalities.

Results are however different when working on migrations data: a much larger number of communities is optimally detected, each covering a small number of contiguous municipalities. This refers to *the Belgian way of life*: households are only willing to move within a small distance, where relatives and other social relations are prominent, and therefore being prepared to commute daily over relative large distances for an appropriate job.

Three common features in these results are worth mentioning as concluding comments.

First, distance still matters nowadays in a small country such as Belgium: although location (and therefore the distance between places) is not part of the method used (Louvain Method), most communities are made out of contiguous municipalities strengthening Tobler's Law in geography, and/or the idea that relationships are more intense between places that are close to each other often admitted in spatial economics (Gaspar and Glaeser, 1998; Fujita and Thisse, 2013).

Second, at the Belgian scale, it is quite unexpected to retrieve provincial borders when dealing with 2011 data, because the origins of the provinces date back to the end of the 18th century, before industrialisation. After decades of advocating Functional Urban Regions and minimising the impact of administrative borders by geographers, provinces are clearly present on all the maps presented in this Chapter. This is especially true for the provinces West-Vlaanderen, Oost-Vlaanderen, Limburg, Liège and Luxembourg that appear to be prominent interaction fields. Note that the province of Antwerp integrates

Waasland, but only for commuting and not for migration communities. The community centred on Brussels includes the province of Brabant wallon and the western part Flemish Brabant, for commuting as well as for migration. Of course these two areas were part of the historical province of Brabant till 1995, but it is worth mentioning that in the eastern part of Brabant (Leuven) forms an independent community. The province of Namur is linked through commuting and migration movements with the area of Charleroi, leaving the western part of Hainaut around Mons and Tournai as a community on its own. We clearly see that there is a juxtaposition of provinces, urban and transportation networks for explaining these maps. The prominence of the provinces in our results is quite unexpected, especially as the political and administrative dimensions of the provinces are losing in importance as spatial units of the organization of economic interaction over the past decades in recent years. In the Flemish Region, it has indeed been agreed that from 2018 most of the provincial "person oriented" services will be transferred to the regional level. In the Walloon Region, the deletion of the provincial level pops up regularly in medias, like a Loch Ness monster. Resistance to change is anyway multifaceted; some institutional competences are still exerted at the provincial level today, explaining probably some of the observed behavioural realities by means of commuting and migration data. Napoleons' decisions still impact current (behavioural) realities!

Third, by taking the case of BCR, the communities go beyond the boundaries of the Region, confirming the sprawling of the metropolitan territory (Thomas *et al.*, 2012). Strong spatial structures emerge from the two databases: the communities anchored in the Brussels-Capital Region extend towards the outskirts (outside BCR), Leuven proves to be an area of interactions in itself, and to the west of the area under study, a spatial reality centred on Asse and Halle emerges. Walloon Brabant presents a geography which depends greatly on the nature of interactions: it forms a whole in terms of commuting to and from the workplace and four distinct areas in terms of mobile telephony (Tubize, Braine-l'Alleud/Nivelles, the *tripole* Wavre/Ottignies/Louvain-la-Neuve and Jodoigne in the east). There are therefore very stable groups of places, and others which are more uncertain, reinforcing certain results published for other cities (Lawlor *et al.*, 2012) and the multiplicity of facets of a city. This therefore calls for great caution and for these realities to be well understood using tools and ad hoc theories, as a consistent policy cannot be pursued if the boundaries observed do not correspond with those in the planning [reinforcing De Maesschalck *et al.* (2015) and Liu *et al.* (2014) for other places and other problems]. We may therefore also question the conclusions of studies limited to Flanders and Wallonia (with or without Brussels).

The method used apply in particular to large databases. We know that today, public and private services have new untapped sources of data which are georeferenced; we may therefore encourage them to equip themselves with observatories and specific tools or to make these data accessible, as they allow an analysis of realities which are otherwise difficult to measure. They constitute a true potential to understand urban processes and to improve planning and governance. Data science and large databases (big-data) constitute new and useful tools for regional and urban analysts, but it is necessary to go beyond cartographic and technological wonder, and to truly understand the new elements they bring in terms of comprehending, measuring and modelling urban complexity. This direction is followed by Chapter 5 that proposes to evaluate the contribution of this kind of new big-data source within a geographical context.

5

Train travel requests: does geography still matter?

Instead of using Census data about observed population movements (Chapter 4), we here use potential movements of people by scrawling data travel requests on the internet. We here limit ourselves to one train schedule-finder application (*iRail*). Travel requests are collected over a 2016 three-months period, and lead to the construction of OD matrices where the flows are the number of requests between two railways stations and hence consisting into potential movement of people. The Louvain Method is applied and detects communities of tightly connected stations on weekdays and week-ends. The language of the request also leads to clear-cut differences about the living spaces (Section 5.3.2).

Once again contiguous stations are more related than distant stations at the exception of large or touristic cities. Results also show the influence of both the urban and rail network structures on the spatial organization of the clusters.

This Chapter is based on works already published in *ISPRS International Journal of Geo-Information* (see Jones *et al.*, 2016)

5.1 Introduction

Despite their limits and critics, information and communication technologies (ICT) are quite important in transportation issues (see Thomopoulos *et al.*, 2015, for a review), and also offer many opportunities in human geography (Kitchin, 2014). Their availability in streaming, that is, the data being continuously updated over time, is especially attractive for transportation planning. In this chapter, we take advantage of an original big-data set: the travel requests by train in Belgium made through *iRail*, a schedule-finder website and application, and we explore the spatial structure of the communities detected in this train travel demand network. The nodes represent the train stations, while the weight of a link (i, j) is the number of travel requests between stations i and j. According to the gravity model of trade (Bergstrand, 1985; Grasland and Beauguitte, 2010) and, in particular, its applications to commuting (see Cervero and Wu, 1997; Wang, 2000, 2001; Sohn, 2005; Mathä and Wintr, 2009; Dujardin, 2001, for Belgium), this travel demand should be proportional to the weight of the origins and that of the destinations. Therefore, the first assumption tested here is that these communities will be consistent with the urban structure of Belgium.

The gravity model of trade, however, also assumes that interactions between two places are inversely proportional to the transport cost or distance between these places (Bergstrand, 1985; Rodrigue *et al.*, 2013). It is important here to note that the railway infrastructure is different from the travel demand network (see Figure 5.1). They share the same nodes (the train stations) but their links differ. For the former, a link represents the presence of a railway line between two stations, while for the latter it is the number of passengers travelling between these stations. Although the railway network is very well developed in Belgium (see Section 5.2.1), its density and operating speed or frequencies vary throughout the country. Literature in transport geography provides various examples of how anisotropy in a transportation network (e.g. high-speed railways) affects the travel times and costs (e.g. Spiekermann and Wegener, 1994, 1996; Vickerman, 1997; Givoni, 2006; Levinson, 2012). Hence, our second hypothesis is that the physical railway network influences (through the travel distance) the shape and size of the communities detected in the travel demand network.

The aim of this chapter is, therefore, to assess the influence of both the urban pattern and railway network structure on the travel demand by train in Belgium. Official data on railway transport use (such as ticket sales) are not available due to privacy rules imposed by the Belgian national railway company; hence, we here use big-data as a proxy, and more precisely the travel requests made on the schedule-finder website and application *iRail* between December 2015 and February 2016.

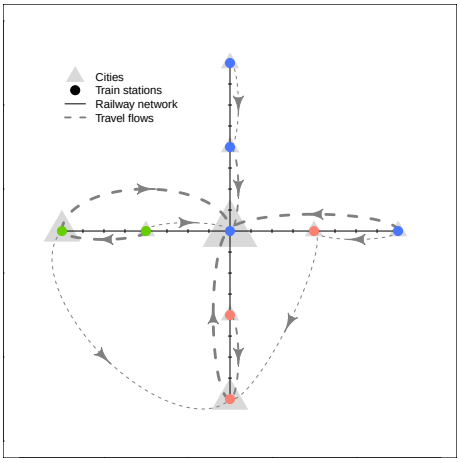


Figure 5.1. Railway network versus travel-demand network (cities and travel flows symbols proportional to their size/volume; the colour of each train station represents its community).

Our methodology consists in applying a community detection algorithm - Louvain Method (Blondel *et al.*, 2008) on this travel-demand network, then to compare the spatial pattern of the communities with the urban and railway network structures, by means of cartographic representation and graph-theory indicators of centrality, following an approach similar to Guimera *et al.* (2005). This procedure is conducted on eight subsets of the data, starting from all requests and afterwards examining temporal, linguistic and regional variations. The chapter is organized as follows. Section 5.2 introduces the dataset, and the methodology used. Results are presented in Section 5.3 and further discussed in Section 5.4 and concluded in Section 5.5.

5.2 Data and methodology

5.2.1 Urban and railway structures of Belgium

Belgium has a clear urban hierarchy (Figure 5.2a) dominated by the metropolitan area of Brussels, centred on the BCR and located in the centre of the country. The BCR accounts for 10% of the Belgian population as of 2016 - 18% for Urban Region of Brussels as defined by Van Hecke *et al.* (2009) - and attracts around 350,000 commuters per day (Thisse and Thomas, 2010a; Verhetsel *et al.*, 2010). Note that although the metropolitan area of Brussels is the largest one in Belgium and includes some neighbouring cities, the city of Leuven stays out of its influence.

All these urban areas are connected by major railroads (Figure 5.2b). In

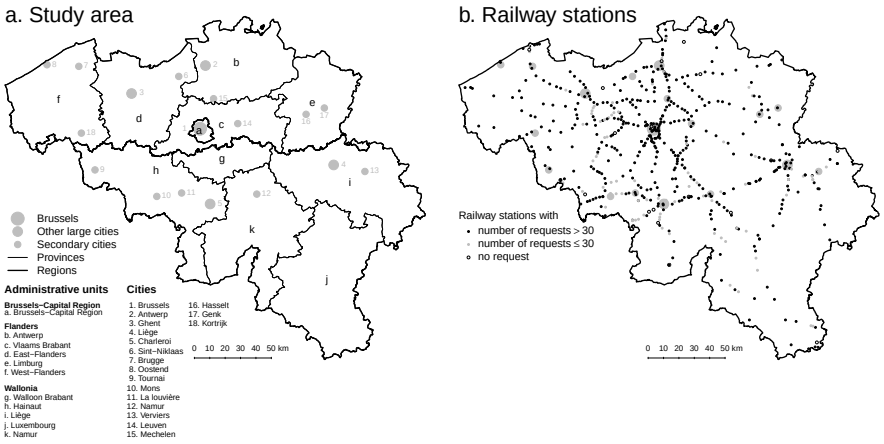


Figure 5.2. (a) Belgium administrative and urban structure, and (b) location of the train stations.

1835, the first railway of the European continent was built in Belgium between Mechelen and Brussels. Many private companies then built their own railways and the maximal extension of the Belgium network exceeded 5,000 kilometres in 1912 (Der Herten, 2001; Denis, J, 1992). Today, the Belgian network is the heritage of major private lines, incorporated into the Belgian National Railway Company (hereafter SNCB[1]). The length of this railways network amounts to 3,595 kilometres (Infrabel, 2019), for a country of 30,528 square kilometres, and is centred on the BCR. In 2015, up to 226.5 million travellers used these railroads - international travels included (SNCB, 2015). Note that an important feature of the network is the so called "Jonction Nord - Midi", linking Brussels - North to Brussels - South stations and allowing the trains to goes trough the city' centre.

5.2.2 Travel demand: the *iRail* data set

Data collection

We collected data from a publicly available webservice ran by *iRail*. *iRail* is both a web-app (`irail.be`) and an API (`api.irail.be`) that allows to query the SNCB trip database, directly from the website, or through a third-party app. See **?** for further technical details. The data on which iRail relies consist

[1]For "Société Nationale des Chemins de Fer Belge". The corresponding Dutch acronym is NMBS. The French abbreviation, however, is more frequent in English and will thus be used throughout this chapter.

of schedules and routes between stations. Information is shared by the SNCB in the standard GTFS format as part of their open data policy. Data for this study have been collected from the API (`http://api.irail.be/logs`), which continuously displays the last 1,000 requests. We therefore polled `api.irail.be` at 30 seconds-intervals to build a comprehensive dataset of the queries. The data collection ran from 20 December 2015 at 10h02 to 28 February 2016 at 20h15, and provided 869,581 queries[2].

Most of the *iRail* API queries consist in requests for trips between two stations, but some include for instance the Liveboard of a particular station. Moreover, in the original dataset, the stations were coded either using a unique identifier, or using a name (which could be in French, Dutch or English). The latter are ambiguous as they only indicated the destination city (e.g. Brussels, Ghent), and not the accurate station, so that we decided to rely only on the railway stations that were identified by a unique id. Finally, we also excluded queries to or from international stations (e.g. Paris North, Lille Europe, or London St Pancras), ending up with an *iRail* data set of 551,301 queries between two Belgian stations.

Each query contains the following fields: timestamp of the query, departure station, arrival station, planned timestamp of the trip (if known), language, user-agent (an indication of the web browser used), and the requesting app. From these fields, we only used the five first. The planned timestamp of the trip was only present in a fraction of the queries, and at the time of the study, could not always be parsed uniquely, so that it was decided not to use it.

Travel requests versus travel flows

The main methodological challenge of the *iRail* dataset is to assess whether it represents the actual travel-demand by train in Belgium. Let us remind that the data collected correspond to the requests made by users online. A request does not imply that a user actually did travel, and vice versa. For instance, users can make the same request several times before travelling, friends of travellers may check on the arrival time of a train, and users who are used to travelling and know the timetable may not need to use *iRail* before their journey. Furthermore, the same information is also available on other websites, especially `www.Belgianrail.be`, the official website of the SNCB.

The question of whether the sample of online requests is representative of

[2]The data collection has been interrupted at the following periods, either due to denial-of-service attacks on the *iRail* server or to crash of our data collection server: 2015-12-31 14:00-17:00; 2015-12-27 05:00-09:00; 2015-12-27 14:00; 2015-12-31 14:00-17:00; 2015-12-31 21:00-2016-01-01 14:00; 2015-01-02 16:35-18:15; 2015-01-03 16:00-19:00; and 2016-01-17 09:10-10:03. To mitigate the effects of a potential failure of our server, we duplicated the data collection, so that no more interruption of the data collection occurred after 17 January 2016.

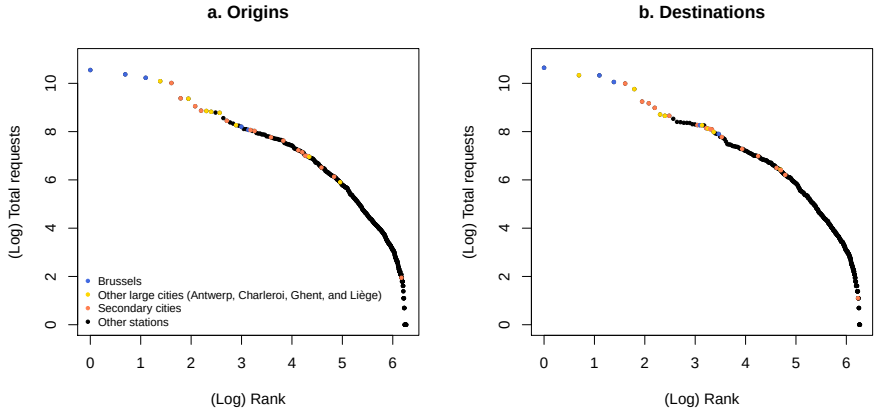


Figure 5.3. Rank - size distribution of the *iRail* requests by (a) origins and (b) destinations.

the actual travel demand is difficult to figure out. Indeed, one may argue that only users who do not already know the timetables will use the *iRail* service, therefore representing only non-routine behaviour. However, the service also offers real time updates of the possible delays, a service which could be used by commuters as well as occasional travellers. Note that, in any case, this chapter does not try to infer absolute numbers of travellers, but rather relies on the relative frequencies of routes among all travels.

Calibration

A comparison of the *iRail* dataset with the observed travel demand is welcome. Since statistics on tickets' sales are not publicly available, this calibration can only rely on the average number of passengers departing from a given station in 2014 and 2015, during weekdays (SNCB, 2014, 2015). The correlation (Pearson) between the log of this variable and the log of the total number of *iRail* requests by origins is high for both years ($\rho = 0.86^{***}$ and 0.84^{***}) and no outlier appears. The number of requests per station in the *iRail* dataset is also consistent with the importance of cities, in terms of rank-size distribution, as shown on Figure 5.3. Note that 17 stations never appear in the requests (25 for origins only, 29 for destinations): all of them correspond to small (rural) stations, and are relatively well spread across the country. Hence, although its limitations must be kept in mind, no strong spatial bias appears in the *iRail* dataset, which can therefore, in our opinion, be used to study the travel-demand by train in Belgium.

5.2.3 Methodology

Subsets of the case study

From the set of *iRail* requests, we create several subsets, to analyse the differences in community detection results, depending on language, temporal or spatial constraints. More specifically, we study (1) the general data set, (2) a subset excluding stations inside the BCR (No Brussel), (3) all requests made during weekdays, (4) all requests made during the weekend, (5) requests made in Dutch, (6) requests made in French, and finally, (7) those made for journeys between stations inside the RER zone. More details about these choices are given in Section 5.3, and characteristics of the networks extracted are given in Table 5.1.

Community detection

Using the data of each of the seven subsets of requests, we create origin destination networks: a station i is linked to another station j with a weight W_{ij} that equals the number of trip request going from i to j or vice versa. On these networks, we apply the Louvain Method for extracting partitions into communities (please see Chapter 2 for further details).

Because the Louvain Method may lead to different results depending on the order in which the nodes are considered, we run the community detection algorithm 1,000 times, and from the results, we extract the dominant structure. To do this, we first detect the cores of communities, that is, groups of nodes that are always assigned together in the same community. Then, we detect the dominant structure, by classifying each of the remaining nodes with the community it is most often assigned to. In this step, two cores of different communities may also be merged if the nodes from both cores are classified together in the majority of the runs.

To compare different partitions of a network, and measure how similar two partitions are, we use the Normalized Mutual Information, or NMI (Ana and Jain, 2003). The NMI measures the amount of information that one partition gives about the other (see Equation 2.6 already proposed in Chapter 2), always takes values between 0 and 1. Zero represents the case where no information is given by one partition about the other, and a value of one the case of identical partitions.

Attributes of the communities

The partitions in communities are further compared using four descriptive statistics. Firstly, for each node we know the number of times that it is affected to its final community over the 1,000 iterations of the Louvain Method (expressed in percentages). These values are presented in Figure 5.9. The

robustness of each community (Table 5.4) is computed as the average of this value over all nodes belonging to that community. We use this stability as a measure of the robustness of the partition obtained.

Secondly, for each journey between two nodes, we can measure the distance travelled alongside the railway network. Note that we did not rely on the travel time, since a travel request made through the *iRail* application may suggest different potential connections, with varying travel times. The shortest travel time may thus change depending upon the time of the day, which is not the case for the travel distance. The mean distance travelled of a community (Table 5.2) is the average of these travel distances (in km) over all requests linking two nodes belonging to the same community.

Thirdly, a spatial contiguity index $SC(i)$ measures if the stations in a community are contiguous in space or, on the contrary, dispersed. For each station i, it is computed as indicated by equation (5.1), where $N_c(i)$ stands for the number of different communities among the station i and its $n-1$ closest neighbours. Note that n is set to five here and that the closest neighbours are determined using the distance alongside the railway network.

$$SC(i) = 1 - \frac{N_c(i)}{n} \tag{5.1}$$

$SC(i)$ varies here from 0 (if all stations belong to a different community, indicating a low spatial contiguity) to $1-1/n$ (when all the n closest neighbours belong to the same community, i.e. 0.8 here since $n = 5$). Average values are then computed for each community, using all nodes belonging to that community (Table 5.3).

Finally, we identify the most central node (i.e. train station) of each community by the betweenness centrality (Freeman, 1977). The betweenness centrality of node i is the number of shortest paths from all nodes to all others that pass through i. It is here based upon the travel demand network, not on the railway network (Figure 5.1). As for the mean travel distance, this indicator is computed using only the journeys between two stations belonging to the same community. Results are presented in Table 5.5. As for the mean travelled distance, the betweenness centrality of each community is computed using only the requests from and to a node belonging to that community. Note that in Section 5.3 we exclude from the results the communities that include less than five train stations. Tables 5.2 to 5.4 are presented in Section 5.3.

Table 5.1. Descriptive statistics of the subsets.

Subsets	**Number of**				**NMI**
	Requests	Stations	O/D Pairs	Com	(with *General* subset)
General	551 301	541	13 770	7	/
No Brussels	265 013	508	9 600	7	0.53
Weekdays	419 184	538	12 262	8	0.79
Weekend	132 117	508	7 374	7	0.58
Dutch	297 710	406	7 473	7	0.38
French	196 185	465	6 927	7	0.32
RER	204 731	166	3 614	5	0.58

Com = communities

Table 5.2. Average travel distance alongside the railway network (km; between bracket = standard deviation).

Subsets	**By communities**								**Mean**
	1	2	3	4	5	6	7	8	
General	32.9 (30.8)	42.2 (27.4)	29.3 (19.1)	32.8 (27.0)	32.7 (20.0)	30.1 (26.0)	46.1 (11.3)		42.8 (32.6)
No Brussels	30.8 (23.1)	41.6 (27.9)	31.1 (21.9)	31.8 (18.5)	9.9 (11.2)	56.9 (38.3)	33.6 (34.7)		44.4 (36.5)
Weekdays	30.5 (31.0)	42.4 (27.6)	33.1 (27.0)	32.6 (19.8)	25.5 (16.3)	30.7 (19.7)	27.6 (24.2)	38.0 (17.8)	42.0 (31.9)
Weekend	33.7 (32.9)	46.7 (31.4)	35.5 (22.0)	26.3 (20.2)	28.1 (19.0)	31.3 (26.2)	41.1 (16.4)		45.3 (34.3)
Dutch	34.4 (21.0)	42.9 (28.7)	20.1 (15.5)	24 (17.3)	26.0 (16.3)	54.6 (36.0)	104.8 (8.4)		43. 8 (31.4)
French	35.9 (25.7)	29.5 (23.9)	31.3 (41.3)	35.3 (27.8)	23.2 (13.1)	39.1 (17.5)	12.4 (7.8)		39.9 (33.3)
RER	14.4 (10.7)	16.6 (11.5)	20.3 (8.3)	28.3 (14.1)	15.1 (5.9)				22.7 (13.0)

Table 5.3. Average value of the spatial contiguity index (between bracket = standard deviation).

Subsets	By communities								Mean
	1	2	3	4	5	6	7	8	
General	0.68 (0.16)	0.61 (0.18)	0.56 (0.18)	0.61 (0.17)	0.56 (0.18)	0.48 (0.13)	0.45 (0.21)		0.61 (0.18)
No Brussels	0.72 (0.12)	0.68 (0.14)	0.66 (0.13)	0.70 (0.13)	0.67 (0.15)	0.51 (0.10)	0.57 (0.07)		0.68 (0.14)
Weekdays	0.67 (0.17)	0.57 (0.18)	0.61 (0.18)	0.52 (0.20)	0.67 (0.16)	0.46 (0.17)	0.47 (0.13)	0.58 (0.21)	0.59 (0.19)
Weekend	0.67 (0.16)	0.65 (0.16)	0.52 (0.17)	0.57 (0.17)	0.53 (0.18)	0.49 (0.14)	0.56 (0.15)		0.60 (0.18)
Dutch	0.54 (0.17)	0.57 (0.18)	0.48 (0.19)	0.55 (0.23)	0.49 (0.16)	0.48 (0.12)	0.50 (0.11)		0.53 (0.18)
French	0.60 (0.16)	0.58 (0.20)	0.63 (0.19)	0.52 (0.13)	0.64 (0.16)	0.62 (0.21)	0.60 (0.23)		0.59 (0.18)
RER	0.60 (0.19)	0.48 (0.18)	0.57 (0.19)	0.55 (0.23)	0.29 (0.11)				0.54 (0.21)

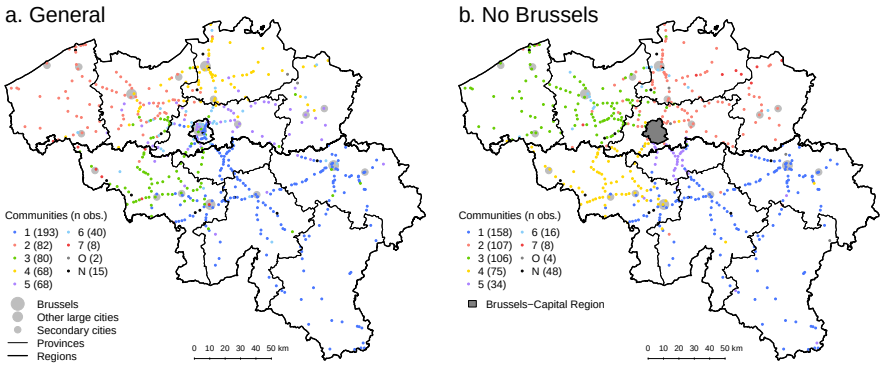


Figure 5.4. Communities in the *iRail* dataset for the (a) *General* and (b) *No Brussels* subsets (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set).

Table 5.4. Robustness of the communities (in %; high values = high robustness; between bracket = standard deviation).

Subsets	**By communities**								**Mean**
	1	2	3	4	5	6	7	8	
General	89.6	98.8	83.3	91.9	87.4	99.5	44.7		90.2
	(14.1)	(3.78)	(19.5)	(14.0)	(9.58)	(2.19)	(0.00)		(14.9)
No Brussels	99.4	88.8	97.8	99.5	96.7	73.6	100		95.9
	(4.11)	(12.1)	(8.10)	(2.25)	(1.58)	(7.13)	(0.00)		(9.22)
Weekdays	96.8	98.9	93.2	96.2	81.1	96.8	91.1	90.5	94.3
	(8.44)	(4.43)	(10.4)	(8.92)	(0.00)	(7.78)	(16.8)	(6.11)	(9.90)
Weekend	97.1	91.9	81.3	79.3	97.4	90.9	88.4		90.8
	(9.49)	(16.9)	(17.7)	(8.21)	(7.94)	(16.0)	(11.5)		(14.6)
Dutch	92.3	94.7	73.5	96.2	97.7	100	98.48		91.8
	(13.5)	(8.06)	(5.06)	(9.10)	(5.93)	(0.00)	(0.00)		(12.2)
French	99.6	95.8	89.1	95.7	100	91.2	94.1		96.0
	(2.90)	(11.1)	(13.9)	(8.87)	(0.00)	(5.39)	(14.3)		(9.28)
RER	97.5	95.3	92.7	74.1	64.5				90.1
	(9.03)	(9.12)	(14.1)	(0.00)	(0.00)				(13.7)

5.3 Results

5.3.1 All requests

Let us first consider all requests included in the *iRail* dataset (descriptive statistics in Table 5.1). The communities detected for this *General* case study are quite stable (Table 5.4), except for the last one (7) that only includes eight train stations. Communities are also highly contiguous in space (Figure 5.4a), which is confirmed by the values of the spatial contiguity index (Table 5.3). The average distances of the requested journeys are similar between most communities, independently of their spatial extent (Table 5.2). Hence, large communities (e.g. 1, 2) do not reflect different travel behaviour, but rather a linkage effect between nodes.

A clear radial structure centred on Brussels emerges. This metropolitan area encompasses the most central nodes of the entire network, and of three communities (3, 5, and 6; see Table 5.5). Note that community 6, organised in a concentric ring around the BCR, constitutes an exception. The central nodes of the remaining communities correspond to major urban areas, although it is worth noting that two of the largest cities (Liège and Charleroi; see Figure 5.2a) are included in community 1 whose central node is Namur, a smaller city but an important crossroad on the railway network.

The *No Brussels* subset (Figure 5.4b) allows studying more in-depth the regional structure of the travel demand, by removing the requests to or from train stations located inside the BCR. Note that the *No Brussels*' subset corresponds to half of the total number of requests and that the similarity with the *General* case is relatively limited according to the NMI indicator (Table 5.1). Both this loss of roughly half the total requests and the limited NMI' value findings highlight the importance of Brussels, compared to other Belgian cities, in the spatial structure of the travel demand by train in Belgium. The robustness is high for all communities of this subset (Table 5.4).

The average travel distances (Table 5.2) are significantly larger for this *No Brussels* subset than for the General one (t = -18.76***). However, it is probably a result of the geographically central position of Brussels in Belgium rather than an indication of a difference in the willingness to travel.

The high values on the contiguity index (Table 5.3) indicate a strong spatial structure of the communities for the *No Brussels* subset. As expected, by removing Brussels from the study area, the radial structure around this city is replaced by a regional polarisation (Flanders versus Wallonia). Flanders is partitioned in two principal communities (2 and 3), corresponding to its western and eastern halves. The centres (see Table 5.5) are respectively Ghent and Antwerp, i.e., the two major cities of Flanders. The spatial structure in Wallonia is also quite strong. Two communities (1 and 4) are close to those observed in the General case study. The province of Walloon Brabant, however,

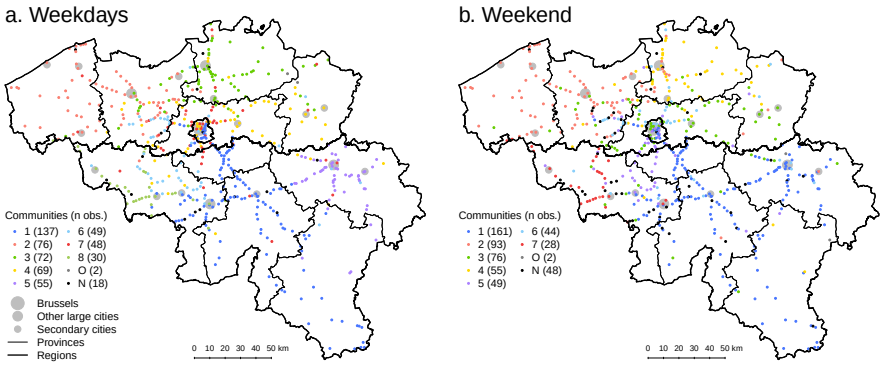


Figure 5.5. Communities detected in the *iRail* dataset during (a) *Weekdays* and (b) *Weekends* (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set).

is here a community in itself (5), separated from community 1.

It is also worth noting that the linguistic border between Flanders (Dutch-speaking) and Wallonia (French-speaking) is well marked: the linguistic border corresponds to the limits of communities; cross border travel demand is small. This is consistent with previous works on home-to-work commuting (Dujardin, 2001) or phone calls (Blondel *et al.*, 2010) in Belgium.

5.3.2 Temporal variations

Travel demand during weekdays is likely to reflect the home-to-work commuting pattern, while the share of leisure trips is probably higher during the weekend. Hence, comparing the *Weekdays* and *Weekend* subsets allows further studying the influence of the urban structure on the spatial pattern of the communities. The former is defined as all requests made from Monday 0am to Friday 12pm, while the latter covers the Saturday 0am to Sunday 12pm period. Given the uncertainties on the time of the actual journey (see Section 5.2), we preferred to separate the requests by days rather than by a finer subdivision.

Although the visual impression (Figure 5.5a) may be different, the communities of the *Weekdays* subset are more similar to the *General* one than those detected during the *Weekend* (see Table 5.1). Let us also note that the weekdays account for 76% of the requests (i.e., more than 5/7), indicating that the number of requests per day is, on average, lower during the weekend.

During the week, the spatial structure of the communities consists of two concentric rings around Brussels, each divided in quadrant. All three communities of the first ring (4, 6, and 7) are close to the BCR, and have their central node located in it (Table 5.5). The second ring is divided into five communities (1, 2, 3, 5, and 8), each centred on a city which is different

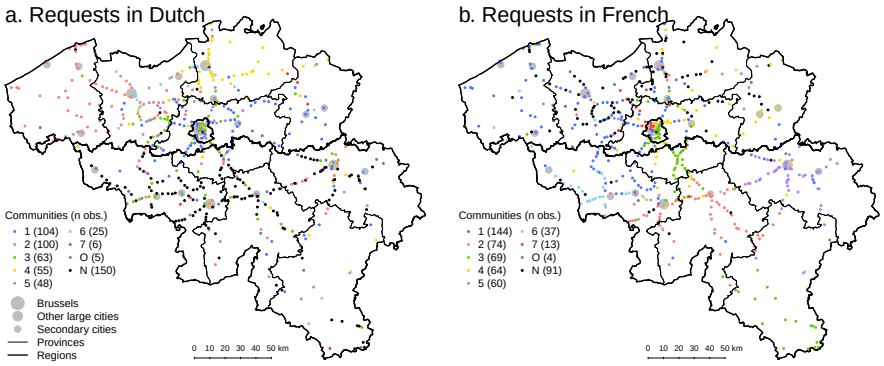


Figure 5.6. Communities in the *iRail* dataset for the requests made in (a) *Dutch* and in (b) *French* (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set).

from Brussels. This second concentric ring seems to gain importance, compared to the General case study, with in particular the emergence of a community centred on Liège and another one on Mons, two regional cities of Wallonia (see Figure 5.2a).

The concentric structure somewhat vanishes during the *Weekend*, replaced by a rather radial one. There are still three communities with a central node in the BCR (3, 5, and 6) and their total importance is similar (169 stations versus 166 for *Weekdays*), although they appear to spread less far away from Brussels (Figure 5.5b). Other communities are very similar to those observed in the *General* case. It should also be noted that the average travel distance of the requests is significantly higher, by about 3 km, during the *Weekend* than during *Weekdays* (t = -30.851***).

Finally, the contiguity of the communities (Table 5.3) is similar between *Weekdays* and *Weekend*, and with the *General* case study.

5.3.3 Linguistic variations

A clear distinction appears between communities located in Flanders and in Wallonia. To further study this distinction, we compare communities detected in the requests made in *Dutch* to those in *French* (Figure 5.6). Both show large differences with the *General* case study (low values of NMI, see Table 5.1), and with each other. The main result is the high number of stations from the other linguistic region (i.e. Wallonia for the requests made in *Dutch* and Flanders for those made in *French*) that do not appear in the subset of the dataset and are thus left black on Figure 5.6. This is especially visible for the *Dutch* subset.

The spatial pattern of the communities in Flanders in this latter subset

exhibits similarities with the one observed for the *General* dataset. Two communities are centred on Brussels (3 and 5), while the central nodes of communities 2 and 4 are Ghent and Antwerp (Table 4). Hence, the spatial structure is not concentric or radial around Brussels, but rather poly nucleated. The train stations located in Wallonia and that appear in the subset of requests made in Dutch do not appear at random: they mainly correspond to major cities (maybe as touristic attraction) and to the scenic Ardennes region. Most are affected to community 3 (centred on Brussels), but no clear pattern appears. Due to this "patchwork" in Wallonia, the contiguity index is very low, both on average and per community (Table 5.3).

This absence of spatial structure in the other linguistic region is more limited for the *French* subset, with most stations in Flanders included into communities 1 or 4 (both centred on Brussels, see Table 5.5). A remarkable result is that community 3 is separated into two parts, the first one in the BCR and the Walloon Brabant province and the second one in the south of the Luxembourg' province. The main reason seems to be the emergence of communities 2 and 5, whose centres are Namur and Liège. As for the *Dutch* subset, the spatial structure of the communities is thus poly nucleated, although Brussels keeps a higher influence.

Despite the lower population densities observed in Wallonia, the average travel distance is significantly smaller ($t = -40.312^{***}$) for the *French* subset than for the *Dutch* one (Table 5.2). This apparent paradox is most likely linked to the lower density (and frequencies) of the train network in Wallonia, and to the tighter urban network in a less hilly landscape in Flanders, and hence to the use of alternative transport modes such as bikes inducing different travelling habits (Verhetsel *et al.*, 2010; Vandenbulcke *et al.*, 2011). The contiguity index is, on the contrary, larger (Table 5.3). Overall, these results show a stronger spatial organisation of the communities in the *French* subset, consistent with the looser urban structure (cities being more isolated from each other than in Flanders).

5.3.4 Regional variations: the RER zone

Brussels emerges clearly in our preceding results as both the city having the largest influence area, located at the main crossroad of the train national network. We here aim at assessing the spatial structure of the travel demand by train in its vicinity, by limiting this time the study area to the RER zone. RER stands for "Réseau Express Régional", a fast suburban train network to and from Brussels, planned by the government and partially completed as of June 2016 (SNCB, 2015; RTBF, 2016). By extension, the RER zone (which was designed by a law, see Moniteur Belge (2004) designates the area where this service operates.

It is clear from the General case study that this zone is divided into different

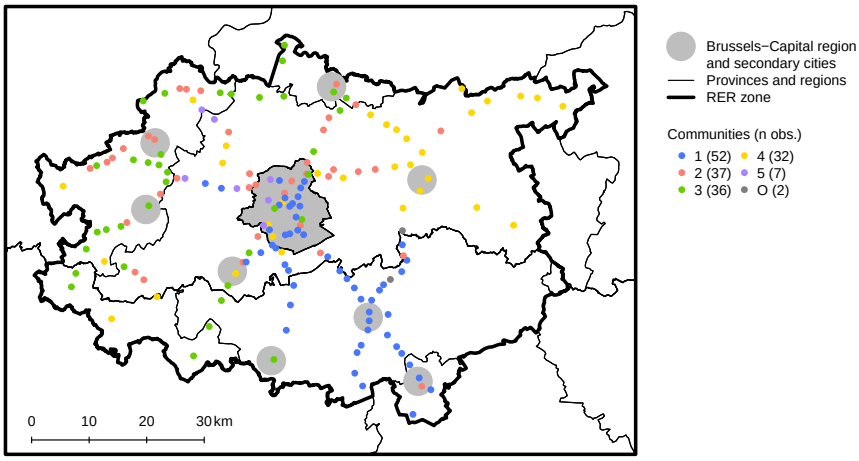


Figure 5.7. Communities detected within the *RER* zone (note: O = Other communities, with less than 5 train stations).

communities, all of them exceeding the boundaries of the RER zone. Note that this RER zone is supposed to encompass all municipalities in *"about 30 km from the Brussels-Capital Region"* (see Moniteur Belge, 2004, pp. 94), which is less than the average travel distance of the requests (42.8 km). One can thus question the relevance of this RER zone delineation, and whether its travel-demand structure is different when isolated from the rest of Belgium. Let us see how far our data helps in answering this question. Figure 5.7 shows the communities detected in the *RER* subset (all requests from and to a train station located inside the RER zone).

Regarding the first question, we already noted that in the *General* subset that only community 6 consisted in a concentric ring around the BCR (Figure 5.4a). Out of its 40 train stations, 28 are located inside the RER zone. Moreover, two of the communities (2 and 4) found inside the RER zone do not even include one station in the BCR. The spatial structure of the communities is, nevertheless, very similar between the *General* and *RER* subsets (Figure 5.4a, Figure 5.7); this is also confirmed by the relatively high value of the NMI in Table 5.1. Note that the RER zone includes 30% of the train stations in Belgium, while the requests from and to these stations account for 37% of the total (Table 5.1), suggesting a high influence of the travel demand inside the RER zone on the communities detected for the General case study.

The average distance between two stations in a request is obviously smaller in the *RER* subset (Table 5.2). The contiguity is also slightly smaller

(Table 5.3), which may be due to the higher weight of the BCR that encompasses a large number of train stations often very close to each other. The main difference is that Brussels is, in the RER case, the central node of all communities except one (Table 4). Hence, their spatial structure is radial around Brussels for the eastern and southern part of the RER zone (communities 1 and 4). The western part is less structured, with two communities (2 and 3) mixed into each other.

Overall, the relevance of the RER zone is not demonstrated in our findings. The influence of Brussels, as measured by the travel demand, exceeds its boundaries towards Wallonia. On the contrary, some areas in the north and west of the RER zone appear to be under the influence of cities located in Flanders. Both results are consistent with the literature on the urban structure of Belgium (Van Hecke *et al.*, 2009).

5.4 Discussion

5.4.1 Urban or network oriented communities?

Our starting assumption was that the spatial structure of the communities reflects both the urban and the rail network structure. A strong influence of the first component should induce a concentric structure centred on metropolitan areas, while the second will lead to a rather radial structure alongside main railway lines. Our results show that both concentric and radial patterns appear in the results (see Section 5.3), although with varying importance across the data subsets considered.

The central node of a community corresponds in most cases to a major city (Table 5.5), and these cities are also the most frequent nodes in the *iRail* dataset, both as origins and as destinations (Figure 5.3). Since the Louvain Method ensures that the communities are composed of train stations that have relatively strong interactions with each other, the communities can, therefore, be considered as "influence areas", similar in their principles to Reilly's market areas (Huff, 1963). Note that the robustness of the community at the node level (Figure 5.9) allows assessing the robustness of these catchment areas.

Belgian cities, however, vary in size (see Section 5.2) and the different data subsets presented in Section 5.3 also suggest a multi-level structure of the travel demand. This is in particular the case for the *No Brussels*, *Weekdays*, and *Weekend* subsets. Brussels, the largest Belgian city, is the only metropolitan area constituting the central node of more than one community, for all subsets (except, obviously, the *No Brussels* case). Antwerp and Ghent (in Flanders) emerge as the centre of one community for most case studies while in Wallonia, a community centred on Liège is also found but for only two of the case studies.

Other results, however, differ from what would be expected of urban

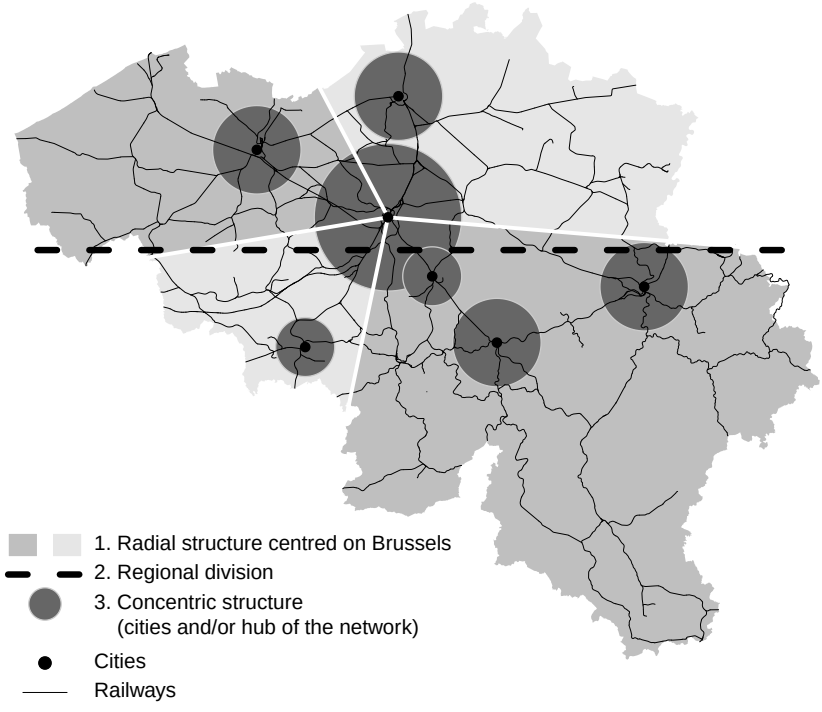


Figure 5.8. Chorematic representation of the travel demand by train in Belgium.

oriented communities. The relationship between the size of the city constituting the central node and the size of the community is tenuous. Among large cities, Charleroi is never the centre of a community. Regarding smaller cities, despite (or due to) the denser urban structure in Flanders (Figure 5.2), no other metropolitan area manages to constitute its own community while in Wallonia the relatively small city of Namur is the centre of the largest community for most case studies.

A radial structure centred on Brussels is visible in most data subsets as well, consistent with a strong influence of the railway network on the communities. This influence seems larger in Wallonia and the eastern part of Flanders, where both the urban structure and railway network are less dense. For instance, the entire Brussels - Namur - Luxembourg line (south-east corner of Belgium) is included in the same community, except for the *French* subset. In the province of Hainaut (south-west corner), we can also observe communities shaped as triangles, with their basis in that province and their summits in the BCR (for *General*, *No Brussels*, and *Weekend* subsets).

Therefore, both the urban and railway network structures influence the spatial pattern of the communities. More precisely, two factors seem to explain the weight of a city in our results: (1) its size (as expected from the classical gravity model of trade), and (2) its position within the railway network. Brussels is both the largest city of Belgium and the heart of its rail network and is, therefore, an outlier. Antwerp, Ghent, and Liège, among other large cities, are main hubs on the network, offering interconnections between several lines. This is less the case for Charleroi[3], the only city that does not emerge as a community centre. On the contrary, hubs of the railway network located in smaller cities (as Namur for the *French* subset), or even purely functional hub (such as Ottignies for the *RER* subset) can still constitute communities' centre. In particular, it is remarkable that for the *RER* subset all stations close to the Ottignies junction belong to the same community, while this is not the case for relatively larger cities such as Aalst, Mechelen, or Termonde. Both factors (the size of a city and its position on the railway network) are consistent with the economic and transportation geography literature, and should be valid for other railway networks than the Belgian one.

Overall, the spatial structure of the travel demand by train in Belgium, as revealed by the *iRail* dataset, consists of three nested layers. First, a radial structure centred on Brussels. Second, a regional division appears, alongside the linguistic border between Flanders and Wallonia. Third, there is a concentric structure around the main hub of the train network, most of them consisting of urban areas. This multi-level organisation is schematized on

[3]To be precise, we refer here to the stations of Antwerp - Central, Liège - Guillemins, Gent - Sint - Pieters, and Charleroi - Sud

Figure 5.8. These findings are consistent with our initial hypothesis that both the urban and the railway network structure will influence the travel demand by train in Belgium. They also show that even in a dense country with relatively efficient public transport such as Belgium, distance is still a key factor of the travel behaviour, as assumed by the gravity model of trade and the related theory of market areas.

5.4.2 Implications for policy decisions

Two main debates exist currently in Belgium on the future of the SNCB. The first one concerns the RER service. Due to budgetary constraints, the completion of the network around Brussels (extension from 2 to 4 tracks per line) is, as of July 2016, uncertain for the Brussels - Nivelles and Brussels - Ottignies lines (RTBF, 2016; Attout, 2016). Our results show that these two lines belong to a community including many stations located inside the BCR, both for the General case and for the *RER* subset (Figure 5.7). Moreover, the five main stations of Brussels[4] are the destination of 18% of the requests from Ottignies, and 80% from Nivelles. This fully supports the importance of completing the infrastructure as originally planned.

A secondary discussion about the RER is that local policy makers from the four other large Belgian cities (Antwerp, Charleroi, Ghent, and Liège) advocates, from time to time, for the development of a RER service in their own city (SNCB, 2015). The analyses provided here, although not specifically designed to answer this question, suggest that providing such service would only make sense in metropolitan areas that are the central node of a community, i.e. Antwerp, Ghent, and Liège, but not Charleroi. Among smaller cities, Namur is also a potential candidate.

The second debate on the future of the national railway company, mostly fuelled by policy makers from Flanders, is its possible separation into two regional (Flanders and Wallonia) companies (Attout, 2016; RTBF, 2016). Our findings do indeed show that most communities do not trespass the linguistic border. This regional division is, however, the second level of the spatial structure described in Section 5.4.1, the main one being a radial structure centred on Brussels. The spatial pattern of most communities centred on that city is independent from the regional boundaries. Moreover, the five main train stations within the BCR are the destination of travel requests from 423 different stations (or 78% of all stations), and the origins of travel requests towards 431 stations (80%). Whatever the future institutional organisation of the Belgian railways, there is thus a clear need to preserve a high level of service to and from Brussels.

[4]Brussels - South, Brussels - Central, Brussels - North, Brussels - Schuman, and Brussels - Luxembourg

Nevertheless, we can question some features of the rail transport offer. In particular, thanks to the "Jonction Nord - Midi" (see Section 5.2) the main stations in Brussels are connected to each other. Different train services can thus call in Brussels while joining other cities, for instance IC-01[5] connection links Oostend to Eupen, via Ghent, Brussels, and Liège (and vice-versa) or the IC-05 connection going from Charleroi to Antwerp, via Brussels. In the *iRail* dataset, 11% of the requests from Ghent-Sint-Pieters have Brussels-Central as destination, while for Liège it is only 0.2%. The corresponding values from Liège-Guillemins are 5% and 0.5%. From the station of Charleroi-Sud, 12% of the requests have Brussels-Central as destination, and 2% to Antwerpen-Centraal. In the other direction, the shares are of 9% for Brussels-Central, and 0.7% for Charleroi. Both examples show that the travel demand between two Belgian cities separated by the linguistic border is low, even if a direct connection exists between these towns. One may argue that a train service from a regional city to Brussels and return would be more efficient. However, assessing this question overcomes the goals of this chapter.

5.4.3 Challenges and paths for future research

The *iRail* dataset offers methodological challenges as well as an important potential for future research. For the former, two main difficulties should be outlined. First, as detailed in Section 5.2, a travel request made on the *iRail* website or application does not mean that the journey was actually made. The number of requests cannot, therefore, be easily translated into forecasts of the absolute number of passengers. Still, we end up with an accurate representation of the relative importance of each link, which is sufficient to study the spatial structure of the travel demand by train in Belgium. The second difficulty is that *iRail* is only one of the various schedule-finder websites and applications existing for Belgian railways. The analysis proposed in this chapter would gain to rely on the travel requests made on the official SNCB application, even if the pattern observed here is consistent with the geography of Belgium.

We were unable to find any published work relying on travel requests to study the travel demand. Nevertheless, datasets similar to *iRail* exist for all schedule-finder websites or applications, the main issue being their limited availability due to privacy or commercial reasons.

Future geographical researches based on datasets similar to the *iRail* one could be conducted in (at least) two directions. The first one is *territorial planning*. The travel requests allow a high level of spatial and temporal details. In a sustainable development context, they could be used to assess the location of a future infrastructure offering the highest willingness to travel by train (or by other public transport means), and the optimal schedules to maximise

[5] "IC" stands for "Intercity", i.e. fast interurban connections.

the modal share of public transport towards this infrastructure. Moreover, ICT data on travels' intentions such as the *iRail* dataset offer information complementary to datasets on observed travel flows. Hence, the second path for research is prospective transportation planning. Let us recall that the analyses conducted here rely on the date of the request, reducing the usability of the data to assess temporal variations of the travel demand. In the context of "smart cities" (Kitchin, 2014), the dataset would benefit from the inclusion in a more robust way of the date and time of the requested journey. A surge of requests towards a given destination (e.g. the Belgian coast) may for instance be used by the SNCB as an indicator that the train's offer should be increased at the time of the requested journeys (rather than relying solely on weather forecasts). Further work is however requested before being able to forecast trains' frequentation in the future based on travel requests.

5.5 Conclusions

This chapter shows that travel requests made through an online application provide a good view of the travel demand by train in Belgium. Using community - detection methods at the station level, we demonstrate the influence of both the urban and railway network structures on the spatial pattern of the travel demand. The results are consistent with the commuting in Belgium. A multi-level structure emerges, consisting of a radial structure centred on Brussels, a regional division between Dutch and French speaking areas of Belgium, and a concentric structure around cities. The importance of secondary cities depends on the temporal, linguistic, or regional subset considered.

The study conducted here shows the relevance of big-datasets for geography and regional sciences, since ICT offer numerous ways to collect (real time) data on travel behaviour. The analyses proposed here also have direct implications for transportation planning, even if further work is needed to ensure their geographical consistency before being able to fully exploit the potential of journeys' requests datasets for prospective transport or territorial planning. Nevertheless, the main result is that the spatial structures emerging from our analyses remain consistent with good old geographical theories, especially the gravity model of trade and Reilly's market areas. In our opinion, big-data should thus not be seen as substitute or competitors of theoretical models, but rather as an opportunity to look at these theories in an original and innovative way.

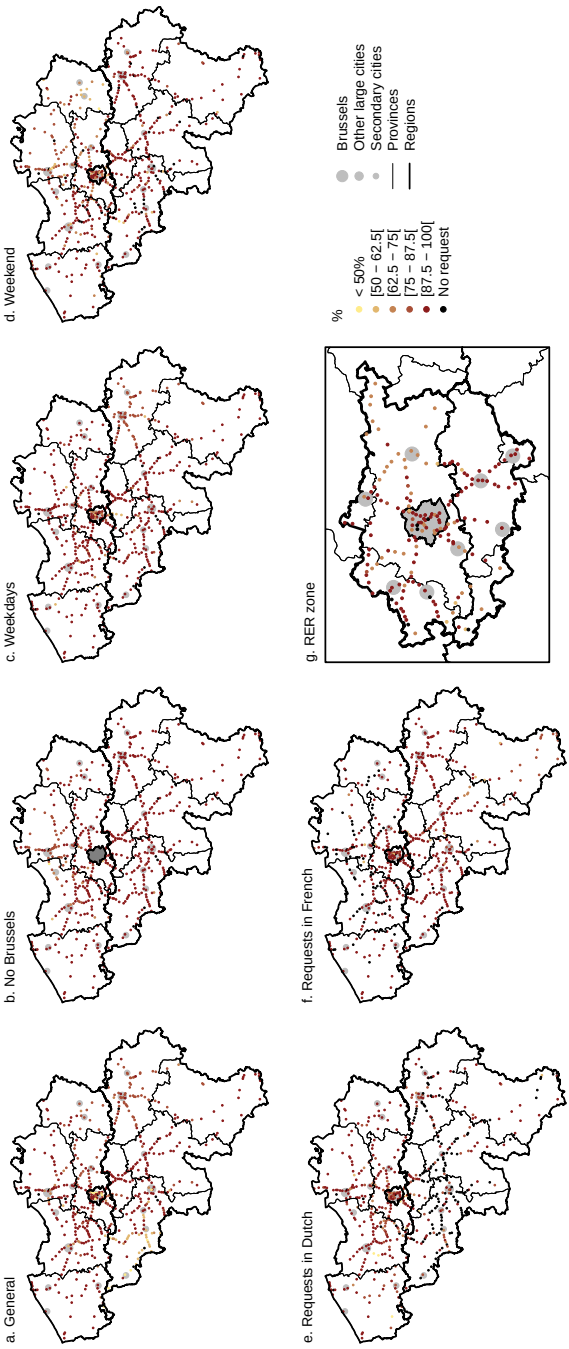


Figure 5.9. robustness of the communities by subset of the *iRail* data set (number of times, over the 1,000 iterations that a node is affected to its final community, in %).

Table 5.5. Central node of the communities by betweenness centrality (between bracket = name of the train station).

Subsets	**By communities**								**Global network**
	1	2	3	4	5	6	7	8	
General	Namur (Namur)	Ghent (Gent-St-Piet.)	Brussels (Bru.-Midi)	Antwerp (Ant.-Centraal)	Brussels (Bru.-Central)	Brussels (Bru.-Nord)	Tournai (Tournai)		Brussels (Bru.-Central)
No Brussels	Namur (Namur)	Ghent (Gent-St-Piet.)	Antwerp (Ant.-Centraal)	Mons (Mons)	Ottignies (Ottignies)	Antwerp (Ant.-Berchem)	Mol (Mol)		Antwerp (Ant.-Centraal)
Weekdays	Namur (Namur)	Ghent (Gent-St-Piet.)	Antwerp (Ant.-Centraal)	Brussels (Bru.-Central)	Liège (Liège-Guil.)	Brussels (Bru.-Midi)	Brussels (Bru.-Nord)	Mons (Mons)	Brussels (Bru.-Central)
Weekend	Namur (Namur)	Ghent (Gent-St-Piet.)	Brussels (Bru.-Central)	Antwerp (Ant.-Centraal)	Brussels (Bru.-Midi)	Brussels (Bru.-Nord)	Mons (Mons)		Brussels (Bru.-Central)
Dutch	Leuven (Leuven)	Ghent (Gent-St-Piet.)	Brussels (Bru.-Midi)	Antwerp (Ant.-Centraal)	Brussels (Bru.-Nord)	Antwerp (Ant.-Berchem)	Zonhoven (Zonhoven)		Ghent (Gent-St-Piet.)
French	Brussels (Bru.-Midi)	Namur (Namur)	Ottignies (Ottignies)	Brussels (Bru.-Nord)	Liège (Liège-Guil.)	Mons (Mons)	Brussels (Simonis)		Brussels (Bru.-Midi)
RER	Ottignies (Ottignies)	Brussels (Bru.-Nord)	Brussels (Bru.-Midi)	Brussels (Bru.-Central)	Brussels (Jette)				Brussels (Bru.-Central)

IV

Movements of freight

6

Exploring freight transportation through the real-time tracking of trucks

Although freight has a huge impact on the economy and on the environment, datasets have always been scarce and limited to small samples. This Chapter tries to overcome this limitation by using the kilometre charge system compulsory for all trucks circulating in Belgium (ViaPass) as a proxy of freight transportation. After cleaning and preprocessing the data, truck traffic is mapped and commented (Section 6.3.1). Particular attention is then put on the geography of the places where trucks stop (Section 6.3.1) before transforming the original raw GPS data into an OD matrix, and further analysed by means of the Louvain Method (Section 6.4). Communities are detected separately for all trucks in Belgium and in the Brussels area, but also in pre-defined categories of trucks based on their country of registration.

Such a large database gives the opportunity to illustrate that transforming a raw data base is a highly time-consuming task that implies many a priori choices. It also leads to the delineation of communities of places highly visited by truck that match well with the geography of Belgium (administrative limits, urban hierarchy, road network, economic activities), as well as with environmental transportation issues at the city scale (Brussels).

This Chapter is based on works already published in a *Core Working Paper* (see Finance *et al.*, 2018)

6.1 Introduction

Since the second half of the twentieth century, the trade liberalisation has resulted in a rise in production and consumption of goods, and hence increasing the importance of the logistic sector (Rodrigue *et al.*, 2016; Verhetsel *et al.*, 2015). Various transportation modes were developed to deliver these goods from the manufactures to the consumption places. They may be grouped into three categories: water, air and land transportation. Each group has its own requirements and features (Rodrigue *et al.*, 2016).

Water transportation corresponds to ships exploiting water routes that cross oceans as well as rivers, canals and lakes. The Belgian waterway network is quite well developed. However, Thomas *et al.* (2003) show that this water network is better developed in the Northern part of the country due to physical constraints (mainly topography) that are less present. The most accessible places reachable with this transportation mode are mainly located in the triangle Antwerp-Ghent-Brussels.

Air transportation, at the opposite of water transportation, has the obvious advantage of speed (Rodrigue *et al.*, 2016) and hence improve the circulation of goods or people on longer distances (ranging between 4,000 km and 12,000 km). There are six airports in Belgium, each having its own characteristics, amongst other internationals, passengers, and freight (Kupfer *et al.*, 2016).

Finally, land transportation mainly includes three main categories: rail, road and pipeline (Rodrigue *et al.*, 2016). The main ability of trains resides and their capacities to transport a large quantity of goods and many people over intermediate distances. High-speed trains have still increased this ability by reducing the duration of the transport. Belgian railways network is dense and well developed (Van Der Herten *et al.*, 2001). It has a polycentric structure following the Belgian urban structure (Vanderstraeten and Van Hecke, 2019) where Brussels dominates. Due to the relative flat topography, the rail network was more densely developed in Flanders and the province of Hainaut rather in the rest of the country (see Chapter 5 and Thomas *et al.*, 2003; Jones *et al.*, 2016). Road transportation of goods corresponds to various kinds of vehicles (cars, vans, trucks, etc.) using the road network. The Belgian road network connects the different urban regions, and is denser in the northern part of the country due to the higher population density. Areas located around Brussels and Antwerp are the most accessible Thomas *et al.* (2003). This Chapter proposes to evaluate only one of these transportation modes: the road transportation made by truck of more than 3.5 tons.

Road transport made by truck has suffered from the lack of spatial data,

restricted to surveys or local counting samples. Currently, this discipline follows a new road of possibilities, thanks to technical and informatics improvements (ICT). These improvements succeed to collect, store, analyse and spread the results linked with the deluge of data newly available from sensors. These objects allow following in real-time people or vehicles, to record their spatio-temporal traces (either this recording is made with the agreement, or not of the user). There is now numerous opportunities to integrate terrestrial or aerial sensors are available (Fan and Bifet, 2013; Kitchin, 2013; Li *et al.*, 2008; Miller, 2017). Although they offer new possibilities, issues related to time-consuming cleaning steps or with the analyses are observed (Goodchild, 2013). The amount of information is so huge that it may scratch and overwrite our techniques and competences. Within the transport geography field, only few empirical works based on exhaustive individual observation (more or less representative) and containing a spatial dimension were available. Commuting and migration movements are probably the only examples of accurate spatial information although the information is often aggregated and suffers from various issues (see Chapter 4 and Commenges, 2013). As regards the transportation of goods, where "more and more data concerning infrastructure, the offer of regular public transport and territories become available" (Dobruszkes, 2012) it is, at the opposite, always difficult to deal with the totality of the transport services, leading to a frustration quite well developed by Dobruszkes (2012).

We here propose a new possibility to deal with spatial analyses of exchanges made by road circulation, thanks to the spatio-temporal tracking in real time of trucks. Because of its exhaustiveness, its use overpass issues linked with surveys implementation and the possible doubt linked with the representativeness. Nevertheless, it questions the quality of the data as well. Is the entire frustration of the unavailability of the data erased by the completeness information of spatio-temporal tracking of trucks circulating within the country?

An accurate understanding of the circulation of trucks between the different cities at the national scale is mandatory to control the various externalities linked with the freight transportation (Macharis and Melo, 2011; Rodrigue *et al.*, 2013). Complementary to these works, we here use this knowledge on the circulation of trucks to better reveal the economic polarisation of Brussels (Merenne-Schoumaker *et al.*, 2015; Strale, 2009). In doing so, we first draw the geography of the transportation of goods by road within Belgium, before looking at the scale of Brussels, where the economic dimension is known has to be centred on logistic, economic and commercial activities that initiate many movements of trucks (Macharis *et al.*, 2014). The objective is to delineate the hinterland of Brussels (analysed from the

circulation of trucks) by revealing one of the dimensions of the economic polarisation of that city. What are the places polarised by the Brussels Region in terms of the movements made by truck? More than just analysing the place of Brussels within Belgium, this contribution aims at exploring spatial traces left by truck registered abroad. In other words, depending on the country where trucks are registered, are the places visited similar? Do trucks registered abroad have the same contribution as Belgian trucks within the national freight exchanges or are the processes drastically different?

This Chapter first presents the data, the various cleaning and preparation steps mandatory to obtain a reliable and well-built spatio-temporal information from a huge amount of raw unstructured data. Second, the transport geography within Belgium and then around Brussels is depicted in order to find the position of the city within the Belgian road transportation system. Third, interactions between places are analysed by the use of the Louvain Method, leading to the determination of freight regions in Belgium. Communities are detected separately for all trucks in Belgium and in the Brussels area, but also in pre-defined categories of trucks based on their country of registration. Finally, a discussion is presented about the pertinence of this newly developed dataset in the objective to better understand the Belgian circulation of goods. All along this Chapter, a particular attention is done on the identification of the classical and new issues emerging, when analysing individual spatio-temporal data which were not initially developed for a scientific purpose.

6.2 From raw GPS data to a spatial information

The key point of this Section is to illustrate the difficulty to obtain relevant and useful spatio-temporal information from a huge raw dataset (Gingerich *et al.*, 2016; Thakur *et al.*, 2015). After a description of the data, Section 6.2.2 explains the methodology used to obtain useful geographical information from a succession of GPS points (in other words, from the GPS trace to a succession of trips and segments). Section 6.2.3 presents the study area and the two delineations developed to reveal the influence of Brussels within the circulation of trucks in Belgium.

6.2.1 Tracking trucks in space and time

Since April 2016, each vehicle of more than 3.5 tons driving within Belgium has to pay a road tax for the use of the 6,500 kilometres of road network. For each truck, the tax is calculated according to the number of kilometres travelled. To do that, trucks have to be equipped with an On Board Unit (OBU) that calculates in real time the value of the tax (http://www.viapass.be/fr/). This

OBU corresponds to a GPS tax tracker sending a GPS point every 30 seconds when the truck is moving.

The positions of the trucks were recorded on each Belgian road for one week, from Monday 14 November 2016 to Sunday 20 November 2016 (and where we decide to exclude the information collected during the weekend). The dataset contains more than 799,000 different ID of trucks, registered in 63 countries and emitting more than 270 million GPS points. Each of these GPS point is characterised by the ID of the Truck, the coordinates of the GPS point, the time (timestamp) when the GPS point was sent, the instant velocity of the truck, the direction of the truck, the country where the truck is registered, the Eurovalue (a pollution class reference in Europe), and finally the MTM that corresponds to the maximum weight that a given truck can carry. This Chapter focuses only on trucks that are "really" driving between places: only trucks that emit more than 10 GPS points per day are considered in this contribution. This choice allows avoiding trucks that are driving on very small distances and that are not particularly relevant in consideration of our objective, like moves within a parking area.

The data used here were initially collected for a taxation purpose and hence not for scientific research. The main objective of this taxation was to internalise the negative externalities (pollution, road damages, congestion, etc.) of the road transportation of goods. After some cleaning and transformation steps, Belgium has at his disposal an unprecedented way to deal with spatial and temporal information on individual scale and on the totality of the trucks' displacement. Either on quantity or quality, the data formerly available (samples from local counting points, surveys, etc) were more limited. Furthermore, local counting points located on specific road segments only allow one-time measurement of the traffic (the flow of trucks circulating on this specific road) without any generalisation on a larger scale. No further information related to the kinds of vehicles, the origin or the destination, as well as the goods transported were available. Similarly, using surveys leads to neglect the exhaustiveness (or worse, the representativeness) of the phenomena due to temporal or financial constraints. Obviously, a sample of logistic companies has to be made, and the low answer rate would inevitably lead to really small samples from where it is difficult to extract generalities (please read Lombard, 1999, in the case of road transportation in France). However, these questions linked with the quality and the quantity of the data is not totally raised in this contribution, as explained in the following Section.

Analysing traces left by objects is not a novel approach: pedestrians, cyclists and cabs have already been studying in spatio-temporal analyses in different geographical contexts and mainly at intra-urban scale (Laurila *et al.*,

2012; Thomopoulos *et al.*, 2015). Other works explore the road freight system using different data sources: sensors were once placed on the road (counting area, surveillance cameras), or installed on board (transponder, tachographs, GPS trackers, smartphones, etc). A synthesis is presented by Antoniou *et al.* (2011). Research focused on real-time tracking are mainly based on data provided by a single GPS system maker, or based on a sample of transport companies. They often answered questions related to the efficiency of the road network (Flaskou *et al.*, 2015) within a specific geographical area - some American states or emerging countries such as China or South Africa (Joubert and Meintjes, 2015; Kuppam *et al.*, 2014; Ma *et al.*, 2011).

The major outline of this literature review is the lack of converging criteria, thresholds or filtering operation to move from a raw GPS data structure into a built, reliable and usable geographical information. The common step within many publications is the identification of the trips and the OD segments within space and time. Shen and Stopher (2014) point out this step as a main "challenge" due to the numerous issues linked with the loss of GPS signals, or the amount of noise recorded in the database. To correctly determine how different trips made by a given truck are organised in space and time, Shen and Stopher (2014) determine various thresholds (temporal or spatio-temporal ones) allowing to correctly split the succession of GPS points into distinct trips. Temporal thresholds varying from a few seconds to minutes are cited without any consensus (Shen and Batty, 2018; Thakur *et al.*, 2015; Zanjani *et al.*, 2015), especially that their significations depend on the geographical context, the objectives of the research, or the data structure (continuing samples or not, variable time gap, recording only when objects are driving, or not, etc.). The low interest of exhaustiveness and representativeness, as well as the absence of validation process are the major critics of these works.

6.2.2 From GPS trace to a succession of trips and segments

Like many databases (big-data does not escape to this important rule), a particular attention has to be done when cleaning. Indeed, several errors could occur in many steps of the collecting process: issues with the OBU (emission of the GPS point after 31 seconds, wrong accuracy or wrong velocity estimation) or during the sample process (wrong matching between the GPS location and the closest road). In this Section, all the cleaning steps are presented, as well as all the choices made to modify the data structure in consideration to our objectives.

Defining the destination of the trucks

The objective is here to transform the large number of GPS points recorded every 30 seconds by the OBU, into a simple OD information (Figure 6.1). Each GPS point is firstly embedded within the spatio-temporal successions of GPS points emitted by a given OBU (the trace). Secondly, each trace is modified to end up into trips (successive GPS points emitted between two stops) and finally leading to OD segments (Origin-Destination summarising trips). The detection of these trips and OD segments release on the identification of the stops made during the movements of trucks. A stop is here simply determined by a temporal criterion: a delay of more of 10 minutes between two successive GPS points. This threshold was chosen to be sufficiently low to catch different nature of stops (loading, unloading, rest, etc.) without integrating small parasite stops (congestion, traffic lights, small OBU distortion, etc.). Based on this 10 minutes threshold, each GPS trace is split into trips for those distance is computed (Euclidean distance between successive GPS points). Trips are simplified into the OD segments that directly linked each couple of OD. The Euclidean distance is computed for each of the OD segments as well.

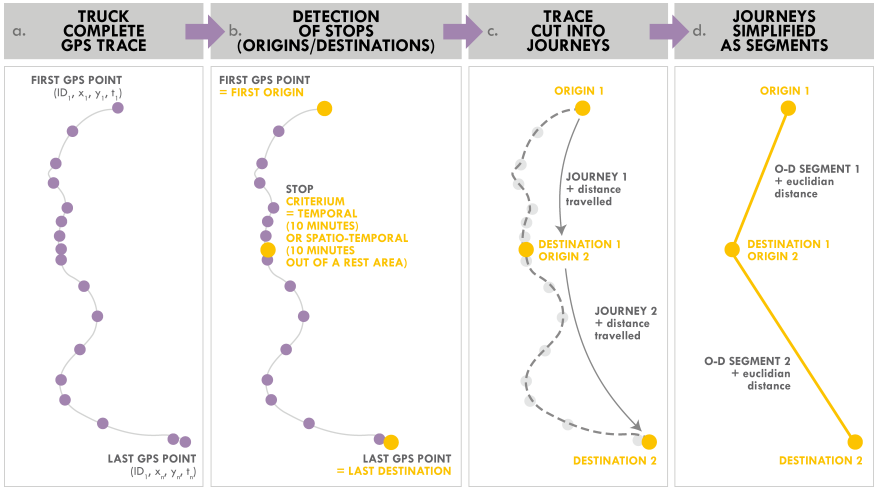


Figure 6.1. Transformation of raw GPS points into trips and Origin-Destination segments.

Sensitivity analyses was made to validate the choice of 10 minutes: Table 6.1 presents some statistics used to detect the optimal value. Analyses were applied in particular values of time ranging from 0 to 300 minutes. Figure 6.2 illustrates the number of journeys detected with the increase of the duration of the stops. Obviously, with the increase of the time threshold, the number of journeys detected decreases. As longer the duration of the stop is,

Table 6.1. Sensitivity analyses of the temporal criterion

Duration of the stop (min)	Time difference (min)	Number of journeys (M)	Lost of journeys (M)	Lost of journeys (M) / time dif.
0	/	15	/	/
0.5	0.5	12.9	-2.1	-4.2
1	0.5	10.07	-2.83	-5.66
1.5	0.5	9.5	-0.57	-1.14
2	0.5	8.42	-1.08	-2.16
2.5	0.5	7.75	-0.67	-1.34
5	2.5	6.3	-1.45	-0.58
10	5	4.45	-1.85	-0.37
15	5	3.58	-0.87	-0.174
20	5	3.05	-0.53	-0.106
30	10	2.45	-0.6	-0.06
40	10	1.97	-0.48	-0.048
50	10	1.74	-0.23	-0.023
60	10	1.57	-0.17	-0.017
70	10	1.44	-0.13	-0.013
80	10	1.36	-0.08	-0.008
90	10	1.3	-0.06	-0.006
100	10	1.25	-0.05	-0.005
200	100	1.05	-0.2	-0.002
300	100	0.99	-0.06	-0.0006

as smaller the loss of journeys is, until reaching only one big journey of 24 hours for each truck. Two indicators may help to determine the best temporal threshold. On the one hand, Figure 6.2 shows that a threshold of 10 minutes corresponds to the beginning of the curvature while a threshold of 20 minutes corresponds to the maximal value. On the other hand, Table 6.1 shows the loss of journeys observed between two time values weighted by the time difference. This expresses the importance of the loss of journeys (and hence corresponding to possible stops). The lost computed when using 5, 10, 15 and 20 may correspond to real stops while the lost observed with values below 5 mainly correspond to some OBU errors (see Section 6.2.2). Given that each truck is different (size, nature of the delivery, working period, etc.), it is reasonable to think that stops are hence specifics of each truck and has to be computed individually. These show the difficulties to find an optimal and general value that could be applied to all trucks.

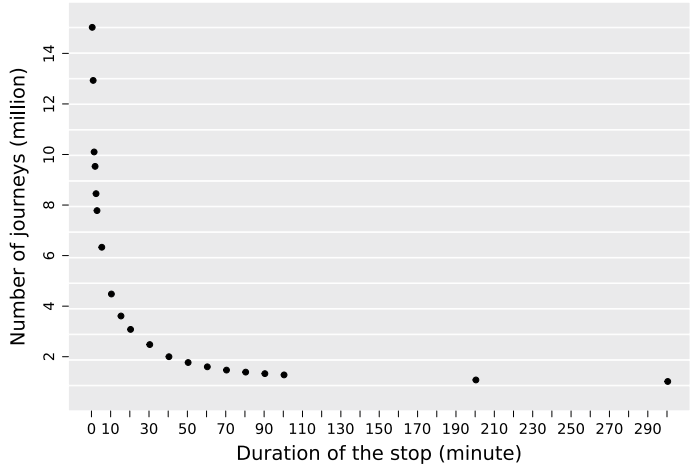


Figure 6.2. Number of journeys detected for different duration of stops.

A daily renumbering of the trucks' ID

Due to privacy issues, IDs of the trucks are randomised each day, making impossible to follow a given truck during the five days of the sample. The renumbering is applied every day at a random time between 01:50am and 02:00am, corresponding to the lowest circulation intensity. Due to that, when the ID of a given truck is randomised, this specific ID disappears from the database and a new one is generated (Figure 6.3), leading to the creation of a wrong destination (last GPS point emitted before the renumbering) and a wrong origin (first GPS point after the initialisation of the new ID). To manage this issue, a new temporal organisation of the database is designed: the entire week is organised in five independent days, from *d* day 02:00am to *d+1* 01:59am. Without this temporal cutting, a truck circulating during the ID's renumbering period would be recorded twice, leading to the creation of wrong ODs. Figure 6.3 shows that between 01:50am and 02:00am, an unusual number of new first Origin and last Destination are made as comparison of the previous or following minutes. Once the new temporal design is applied, IDs appearing only after 01:50am are excluded from the analysis (probably the continuum trips of a truck renumbered), and the OD recorded between 01:50:00am and 02:00:30am are labelled as fuzzy (probably the last/first GPS point before/after the renumbering).

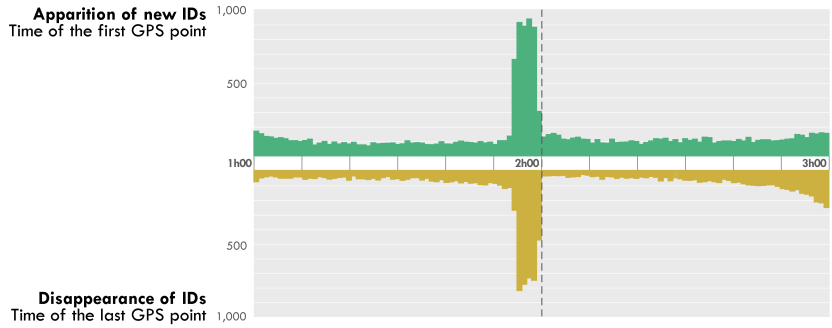


Figure 6.3. Appearing and disappearing of the trucks' ID around 02:00.

Validation and correction of the traces, trips and segments

This paragraph presents the corrections made, as well as the validation choices applied on the database, once the restructuring is finalised. These steps are made possible by the comparison of the network distance and the Euclidean distance, as well as the timestamp and the velocity observed between couples of GPS points and between couples of OD. The three following points summarise corrections and suppression made when inconsistencies were identified (Figure 6.4).

First, when the time is short between three successive GPS points (no stop detected) but the distance is great (more than an arbitrary threshold fixed at 3 km), the inconsistent intermediate point is erased and new distances are recalculated (Figure 6.4.a). A bad reception of the GPS signal or a wrong transmission of the information by the OBU is the possible explanations for this kind of exceptional error. On the opposite, if this error is recurrently observed for a given ID, the totality of the trace is erased from the database (the renumbering of the ID is made by the OBU itself and it is possible that one given ID is used by two different trucks circulating at the same time).

Second, when there is a small distance between a detected destination and the following origin (arbitrary threshold fixed at 3 km), but these two points are found in two different BSUs (meshes of 1 km^2 - grid is presented in Section 6.2.3), points are attributed to a given mesh. This given mesh is chosen as regard the instant velocities measured by the OBU at the OD, when considering that the most accurate location is characterised by a velocity equals 0. This error is mainly due to the slow reactivation of the OBU after the departure of the truck. The correction is applied as many time as the inconsistency is observed.

Third, when the distance between the destination and the next origin is

consequent (more than an arbitrary threshold fixed at 3 km) although a stop is detected, the reliability of the OD is questionable (Figure 6.4.c). Indeed, this situation does not correspond to a stop but reveals an issue within the trace due to a probable loss of the GPS signal by the OBU, an error during the uploading of the data, an internal error of the OBU or a (voluntary) breakdown of the collecting system. A correction is applied only when this error is observed once in the trace of the truck. If it appears several times, it is impossible to rely on the veracity of the stops detected. To correct this issue, the trips and OD segments that include the stop are merged, leading to the creation of new OD points.

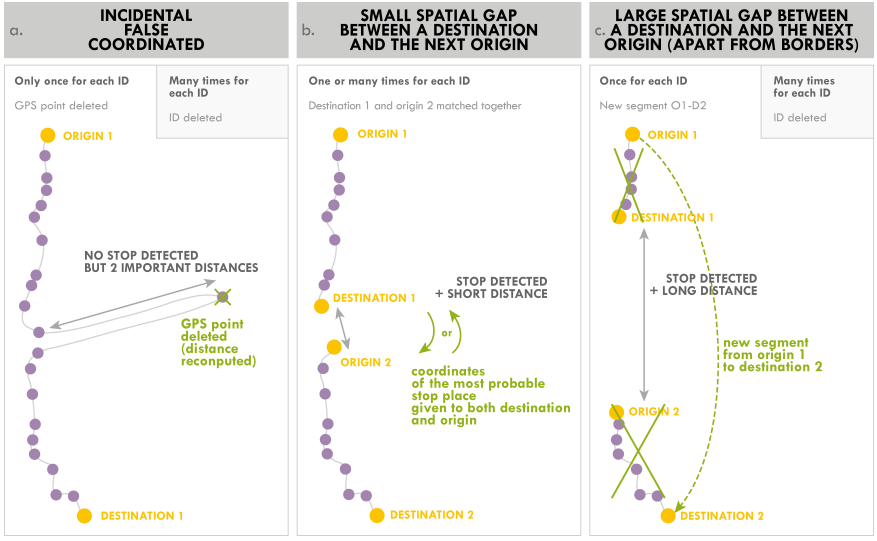


Figure 6.4. Main steps of the detection and correction of inconsistencies, according to the occurrence of errors observed for each ID.

At the end of this cleaning process, we ended up with the suppression of 11% of the GPS points and 21% of the truck's ID from the initial database (Table 6.2). This process shows the importance of cleaning the data in order to manage inconsistencies drowned within the amount of information. It remains important to stay critical in comparison to the quality of the data collected by sensors.

Table 6.2. Retained information after the main filtering, cleaning and validation steps.

	Trucks'ID		GPS points		OD segments	
	N	%	N	%	N	%
Initial Database	798,453	100.0	269,194,440	100.0		
After erasing the trucks' ID with fewer than 10 GPS points	719,425	90.1	265,564,849	98.6		
After the temporal reconstruction	655,701	82.1	253,684,428	94.2	3,746,263	100.0
Final database after the main steps illustrated in Figure 3	628,769	78.7	238,927,089	88.8	2,758,940	73.6

6.2.3 Delineation of the study area

After first analyses centred on Belgium (Section 6.3.1), we propose to concentrate our observations on Brussels and its importance in the Belgian freight network. To do so, the entire country is divided into a grid composed of meshes of 1 km^2 (approximately 32,000 meshes). This grid was defined by Eurostat (2016) to map the spatial distribution of the European population. Two definitions of the study area are chosen to conduct our analyses.

On the one hand, the delineation of the "Brussels area" is based on the methodology developed by Vanderstraeten and Van Hecke (2019) and corresponds to the Brussels Agglomeration. Cells that are included within the Brussels Agglomeration are here considered as the first definition of Brussels (and called "Brussels Agglomeration"). This definition is used for the analyses of the traffic and connection (see Figure 6.7).

On the other hand, each cell included within the Former province of Brabant correspond to the second delineation used in this contribution (called the "Metropolitan area of Brussels"). This second definition of the study area consider Brussels as well as some major cities closely located (Leuven, Wavre, Nivelles, Hal, etc.). This second delineation is used for the community detection analyses (see Figure 6.12).

6.3 Brussels within the network of trucks movements

This Section aims to explore the circulation of trucks in Belgium, to reveal how Brussels is embedded in the global network of truck movements. By analysing this circulation network at the global scale (Section 6.3.1), and by marking off the trips made from/to Brussels (Section 6.3.2), we here approximate the delineation of the area economically polarised by Brussels.

6.3.1 Circulation of trucks at the scale of Belgium and Brussels

Analysing Brussels through the circulation of trucks requires to understand the processes appearing on a national scale. The circulation of the trucks are decomposed into two main components: the *traffic* and the *connections*.

The traffic

First, we define the *traffic* in order to visualise the disparities within the intensities of trucks moving within the whole territory. Instead of just mapping the number of GPS points recorded in each mesh (that would be highly dependent on the specificity of the roads - velocity, curvature - or temporal events like congestion), the number of different trucks ID in movement within each mesh (hour per hour) is calculated. Hence, trucks that are moving slowly are not considered (because these trucks emit more GPS points than trucks moving quicker in a given period of time). The time information extracted from the entire trace is summed to 24 hours in order to obtain a daily traffic. There are then averaged on the five working days, leading to a daily averaged traffic. Figure 6.5 clearly reveals a strong hierarchy of the road network, which presents high intensity of traffic, mainly around the largest cities. The ring roads around Brussels and Antwerp are particularly well distinguishable. Motorways linking the major economic clusters between them and to the neighbours' countries are the places where the traffic is the most intense. Here appears one of the indirect advantages of the Viapass System: the possibility of an accurate measure of the intensity of the traffic for each point of the territory and in real-time. The Viapass System avoids using a dedicated infrastructure (local counting points), leading to extrapolate information to roads not covered by samples. At short-term, the exploitation of the Viapass System allows authorities to propose ways for improving the road network management, and to detect in real-time unusual congested areas that could be solved by active proposals (traffic diversions, road closures, direct reaction of emergency services).

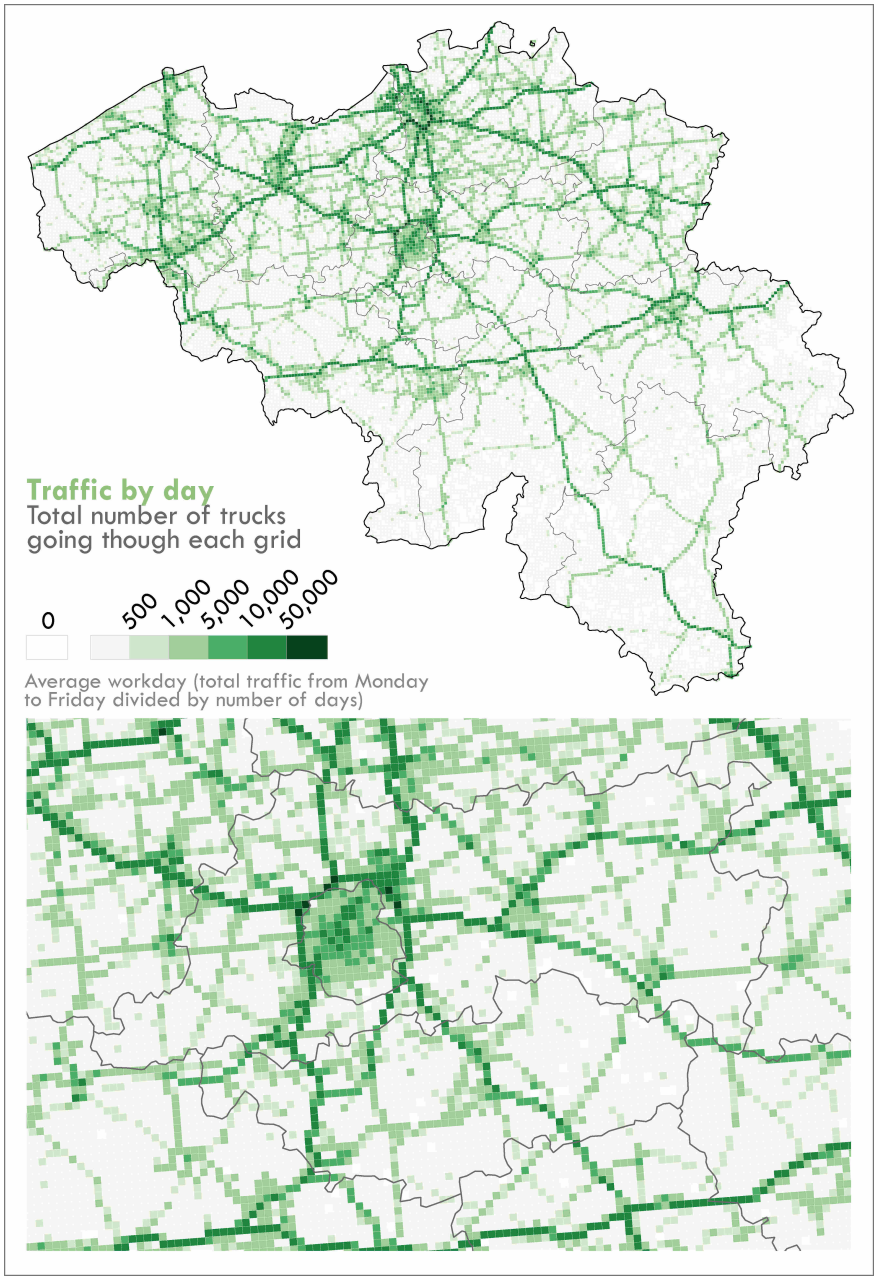


Figure 6.5. Trucks in Belgium: average daily *traffic*.

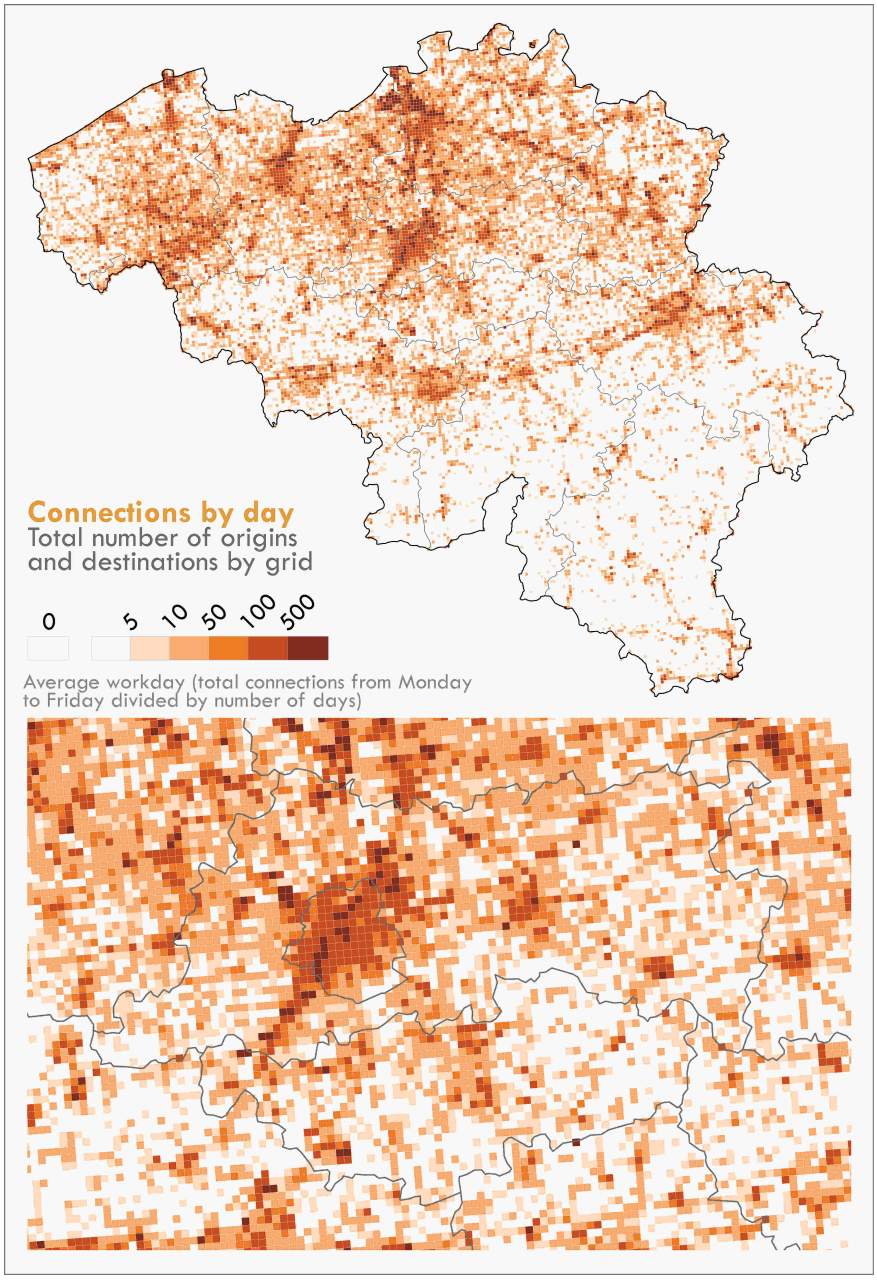


Figure 6.6. Trucks in Belgium: daily averaged *connections*.

Mapping places where trucks are passing through, does not reveal the places really visited by truck, which is measured in the next Section.

The connections

Beyond the visualisation of where trucks are passing through, this contribution aims at detecting the *connections*, that is to say the sum of the number of OD per mesh. OD are here extracted independently of the roads followed. Hence, a truck emitting GPS points in a given cell without making a stop, is not recorded in our database. Similarly to the *traffic*, *connections* are initially constructed hour per hour, cumulated on the entire day, and averaged on the week (five working days), leading to an intensity of daily averaged connection (Figure 6.6). With this second approach, places of economic activities based on potential exchanges of goods are revealed. As expected, major urban poles, (air)ports, and industries are clearly put forward. High intensities of connection are found close to Brussels and more especially within the North-East - South-West axis, following the Brussels-Charleroi canal, or around the Zaventem airport, fitting well with the structure already explained by Macharis *et al.* (2014). Other places are clearly detectable, like 1) ports around the cities of Antwerp, Ghent and Zeebruge or 2) industrial axis Antwerp-Hasselt (along the Albert Canal and the motorway E313), or 3) Liège-Namur-Charleroi (along the Sambre-et-Meuse sillon and the motorway E42). *Connection* levels have a strong hierarchy and both approaches (*traffic* and *connection*) reveal distinct realities with a complementary message.

Hence, in Brussels, although the *traffic* is intense on the external ring road and on roads radially organised around the city, *connections* mostly correspond to places located close to the canal (North-East - South-West axis) where many economic activities, industries, logistic centre or commercial units are located (Macharis *et al.*, 2014). Split entire GPS traces of trucks into OD segments hence reveals places from where trucks are coming and to where they are going. Unfortunately, the database does not contain any information related to the loading or unloading of trucks, neither in quantity nor nature.

6.3.2 Revealing the economic polarisation of Brussels by the circulation of trucks

Few methods are proposed to identify the hinterland of Brussels in terms of exchanges of goods, approximated by the circulation of trucks in Belgium. This Section analyses the spatial footprint of the hinterland of Brussels by selecting only the trips with an origin or a destination in the Brussels agglomeration (first study area delineated - see Section 6.2.3), and then applied the traffic and connection approaches developed in the previous Section.

Traffic between the Brussels agglomeration and Belgium

The absolute daily traffic from or to the Brussels urban agglomeration is first compared to the relative daily *traffic* from or to the same area (Figure 6.7). Figure 6.7.a is similar to the national *traffic* but it is only measured with trips having the Brussels agglomeration as origin or destination. Figure 6.7.b compares the absolute daily *traffic* into the national *traffic* observed in each mesh in order to build the proportion of *traffic* linked with the Brussels agglomeration (in other words, the probability than a given truck circulating within a mesh is going to, or coming from the Brussels agglomeration). Figure 6.7 shows that the *traffic* of trucks having an OD in the Brussels agglomeration is consistent with the *traffic* observed at the national scale. Nevertheless, a few observations can be made. Main roads and highways close to the area of Brussels are clearly observed, but the *traffic* close to the border of Belgium is not so important. Trucks linked with the Brussels agglomeration have a higher propensity to stay close to the study area and use the highways to join Belgian places located further away (Figure 6.7.a). This observation is confirmed by Figure 6.7.b and the illustration of the relative daily *traffic*. The notion of proximity, and hence distance, is an important variable for the logistic sector: most of the trucks linked with the Brussels agglomeration circulate close to the city, without considering the type of roads and the direction of the other main Belgian cities (Antwerp, Ghent, Liège or Charleroi). It is quite surprising to see that trucks having an OD in the Brussels agglomeration are slightly linked with neighbouring countries. We could believe that the Airport of Zaventem is a main entry gate within Europe due to its central place, but other locations, such as the airport located in Liège or the Flemish harbours have a higher attractiveness. The freight place of the Zaventem airport is probably to be found at the national distribution scale (Macharis *et al.*, 2014).

Connections between the Brussels agglomeration and Belgium

The comparison between the absolute *connections* from, or to the Brussels agglomeration with the relative daily *connections* from, or to the same area is mapped in Figure 6.8. On the one hand, Figure 6.8.a corresponds to the number of segments having the Brussels agglomeration has an origin or destination while, on the other hand, Figure 6.8.b presents the same information, but within the connections made at the national scale. This second information hence corresponds to the probability that a truck making an OD in a given mesh is coming from, or going to the Brussels agglomeration. Figure 6.8 presents places visited by truck coming from or going to the Brussels agglomeration. Stops of the trucks are mainly concentrated around the Brussels agglomeration or at higher distance close to other urban centres, industrial areas, logistic

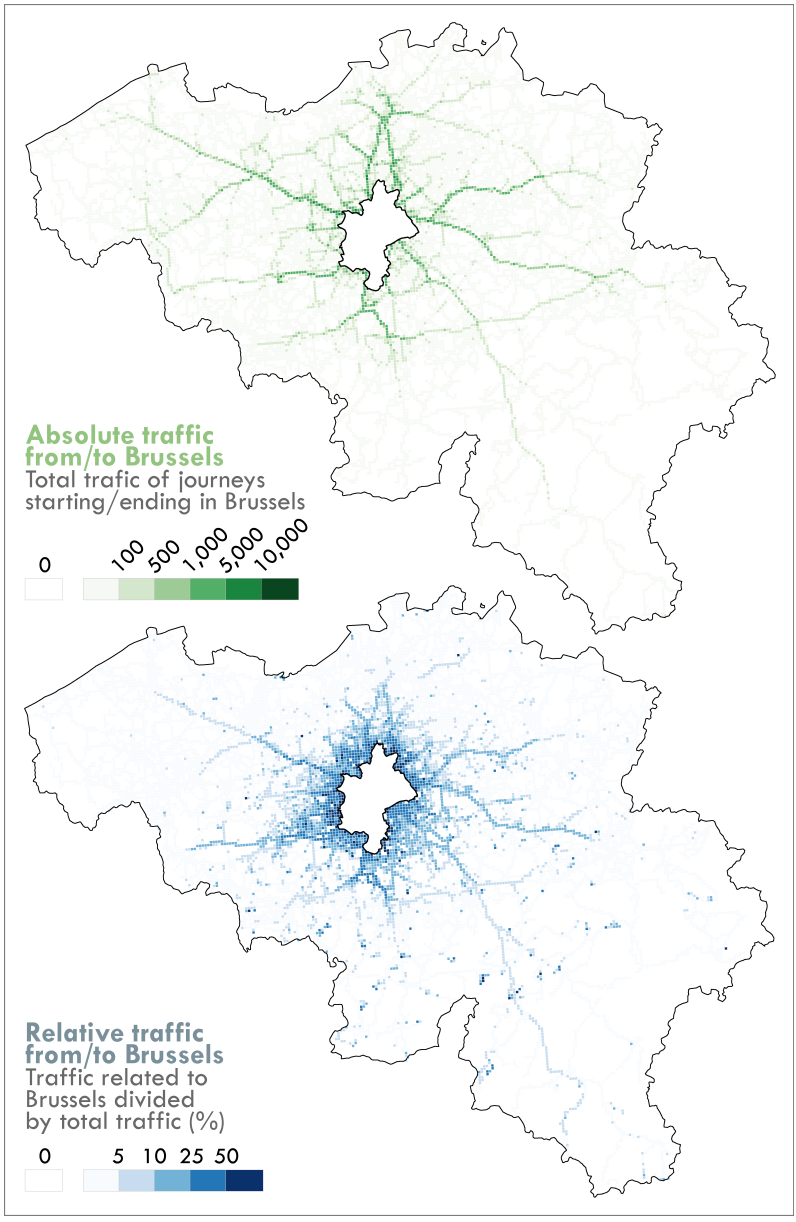


Figure 6.7. Absolute daily traffic (a) and relative daily traffic (b) of trucks having the Brussels agglomeration as Origin or Destination.

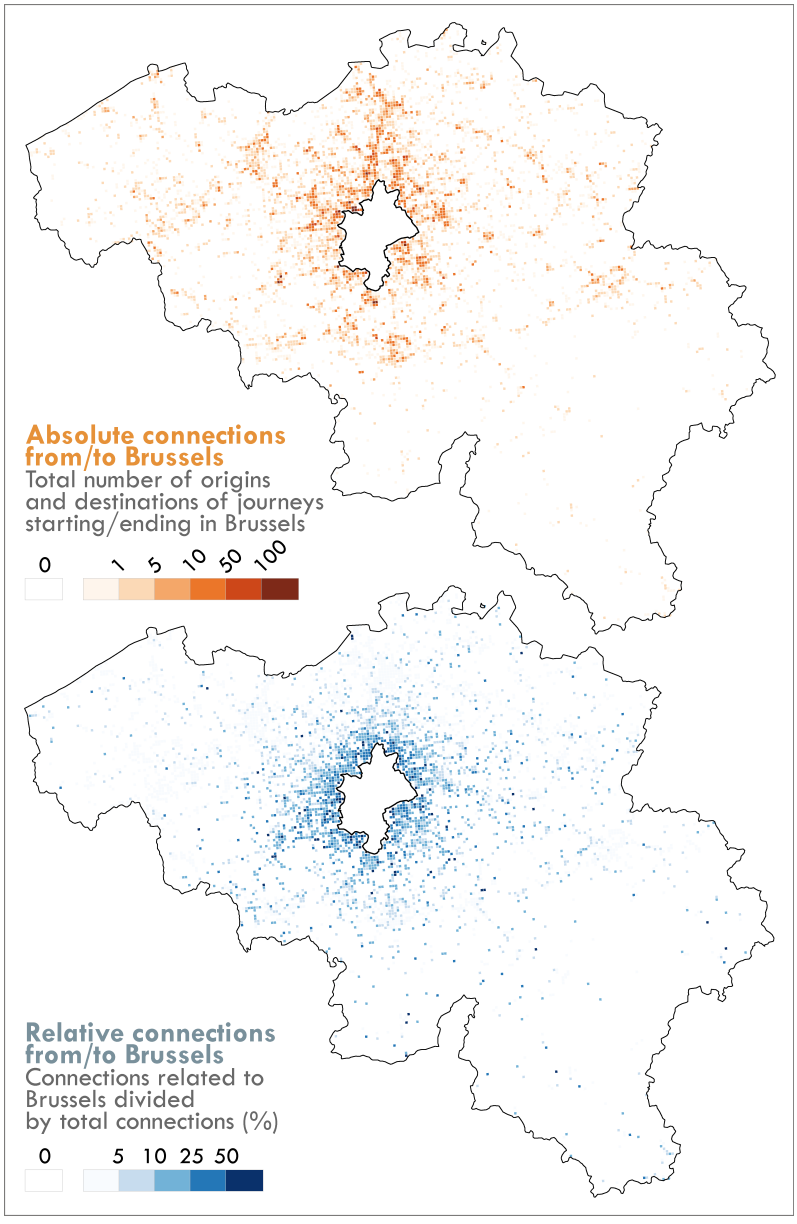


Figure 6.8. Absolute OD (a) and relative (b) of trucks having the Brussels agglomeration as Origin or Destination.

centres and (air)ports (Figure 6.8.a). As expected, the axis Brussels-Antwerp, the Albert Canal and the Sambre-et-Meuse, as well as places close to Ghent, Antwerp, Charleroi and Liège are locations that are highly connected with the Brussels agglomeration. As a complementary information of the absolute information, the relative *connections* (Figure 6.8.b) show that the most intense *connections* are located close to the Brussels agglomeration. Places that are highly connected within the absolute *connections* are not found within the relative ones. Obviously, at the opposite of meshes close to Brussels, these places are not making the totality of their exchanges with our study area and are hence connected to other places.

6.4 Brussels within the (inter)national circulation of trucks

This Section explores the movements of trucks made in Belgium in order to reveal the extension of the area polarised by Brussels when evaluating segments between pairs of places. By analysing this circulation at the global scale (Section 6.3.1), and by marking off the trips made from/to Brussels (Section 6.3.2), we here approximate the delineation of the area economically polarised by Brussels.

6.4.1 Using different networks to catch different realities

Belgium is a small country but an important well-connected crossroad with a highly densely populated region in Europe, which led to many international exchanges. This Section aims to understand how Belgium and Brussels are organised when considering the country of registration of each truck. To answer this question, the nationality of the trucks circulating in Belgium has to be demonstrated. We hence develop two ways: 1) based on the country of registration, we create two categories (trucks registered in Belgium or abroad) and 2) based on the analysis of the location of the GPS points (for a given truck, if only one GPS point is located out of the Belgian territory, we hence consider this truck as an international truck, noted *inter* truck, and at the opposite, if the truck always stays within Belgium, we consider it as a truck circulating exclusively in Belgium, and noted *intra* truck). We end up with four categories: 1) *Belgian* and 2) *foreign* trucks based on the country of registration, and 3) *inter* and 4) *intra* measured by the location of GPS point.

Figure 6.9 presents the proportions of GPS points emitted by countries of registration and their evolution in time. A clear and regular pattern is observed from Monday to Friday. More than 60% of the GPS points emitted within Belgium are made by trucks registered in Belgium. Netherlands, Poland and Germany are respectively the second, third and fourth countries highly

represented in terms of GPS points. Nevertheless, during the weekend, less than 50% of the GPS points are emitted by the *Belgian* trucks and the importance of the foreign countries increases. This reduction of emissions of *Belgian* trucks is obviously linked with the economic activity. Many activities are closed during the weekend or the intensity of the work is reduced (manufacture, shopping centres, etc.). At the same time, *foreign* trucks are probably covering larger distances to join important logistic hubs spread around Europe (ports, airports, highly specialised manufactures).

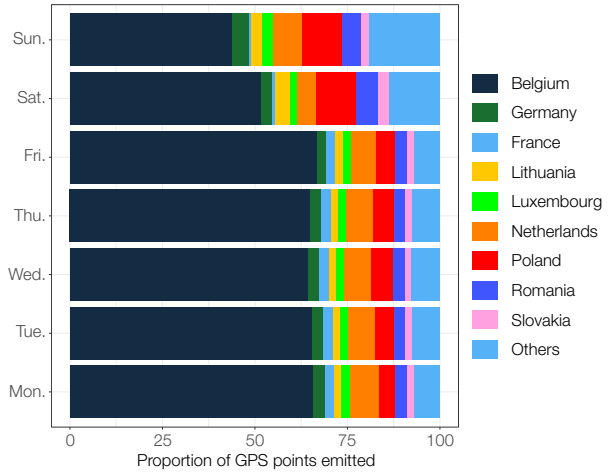


Figure 6.9. Proportion of GPS points emitted by countries of registration by day.

Figure 6.10 presents the daily evolution of the number of GPS points emitted and the number of trucks in circulation for the four categories developed. These two variables follow an identical pattern: a high intensity between 4am and 6pm, with a reduction initiated around 2pm. The four different categories accurately follow two main tendencies. As regards the number of GPS points emitted or the number of trucks in circulation, *intra* trucks and *Belgian* trucks have a similar shape, although *inter* trucks looks alike *foreign* trucks. For more detailed information, the *intra* category is composed of 92% of trucks registered in Belgium while the *inter* category is only constituted by 18% of *Belgian* trucks. Within these *Belgian* trucks, 11% have a first origin and a final destination within Belgium. This expresses that only a few *Belgian* trucks (7%) are leaving the country for more than one day.

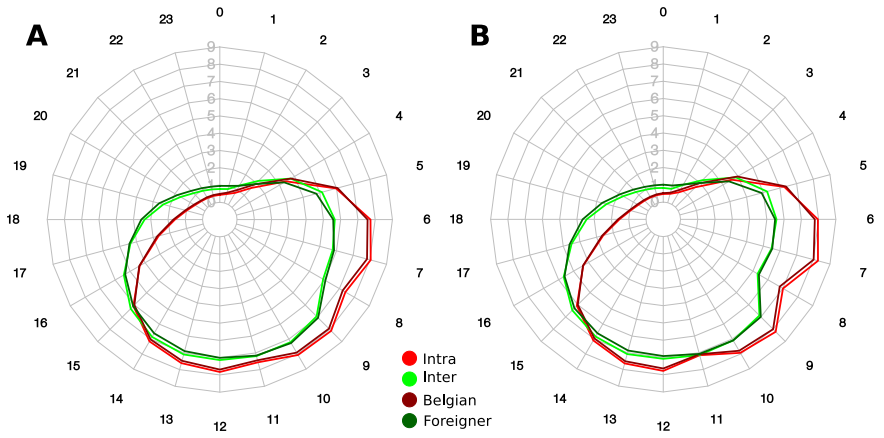


Figure 6.10. Daily evolution of the proportion of A) the number of trucks per hour and B) the number of GPS points emitted per hour, for the four defined categories of trucks.

6.4.2 Determining communities of movements

The place of Brussels on a global scale

Figure 6.11 presents the communities detected by the Louvain Method (see Chapter 2 for details about the method) within the four different categories of trucks developed in Section 6.4.1. Communities mapped in Figure 6.11.a correspond to our benchmark and reveals a particular issue. Dealing with a general network leads to evaluate each linkage between places and hence considering sporadic random events (for instance lunch time stop in the countryside) as an important event in the network. It was thus decided to erase places with a small number of OD: meshes visited fewer than 10 times are hence erased from each network, leading to detect more robust communities. Obviously, the existence of linkages of really small intensity, leads to connect groups of places that are not linked in the "cleaned" networks. Figure 6.11.b illustrates communities found in the "cleaned" general network (that only consider places intensively visited), as regards for the benchmark (Figure 6.11.a).

Figure 6.11.C shows communities found in the "cleaned" *intra* trucks network while the Figure 6.11.D corresponds to the analyse of the "cleaned" *inter* trucks network. The last two illustrations of the Figure 6.11 present communities found in the "cleaned" network of *Belgian* (Figure 6.11.E) and *foreign* (Figure 6.11.F) trucks. It is important to specify that each couple of maps (C-D and E-F) are in fact the networks embedded in the general "cleaned" network. In other words, the addition of these couple of maps leads

to the Figure 6.11.B.

The spatial organisation of communities found in the general "cleaned" network (Figure 6.11.b) is made of large groups of contiguous places that are highly interconnected and hence "exchanging" many trucks. These large groups correspond to the area of influence of the main central economical and logistics Belgian centres, fitting well with the urban hierarchy (Vanderstraeten and Van Hecke, 2019). The linguistic border is quite well followed, with the exception of places located between the Walloon Brabant and the BCR. In addition, provincial boundaries are also well followed, probably due to the provincial organisation of transportation actors.

This spatial structure is similar to the ones in Figures 6.11.c and 6.11.E.: *intra* trucks and *Belgian* trucks have the same partition of space, with the exception of the area squeezed between the city of Tournai and the border of France. These two categories of trucks lead to obtain a similar partition of space.

By analysing the Figures 6.11.d and 6.11.f, similarities are found in communities detected in the *inter* and the *foreign* trucks. Nevertheless, spatial organisation is not similar to the partitions mapped in Figures 6.11.c and 6.11.e: instead of a regional polarisation centred on economical poles and cities, the road network is particularly clear. Stops made by *inter* trucks and *foreign* trucks are mainly organised around the major transportation axis. Places visited by these two categories of trucks have to be easily and quickly reachable from neighbour countries. Communities show that some trucks are only driving through Belgium without making any stops (or within rest areas close to the main roads). For instance, the green community shows that places located close to the German and the French borders (at two opposite sides of Belgium) have high interactions. These places correspond to entries and leaving areas of trucks that are only passing through Belgium. Finally, it is important to notice that the area of Brussels does not correspond to a specific community (Figures 6.11.d and 6.11.f). Brussels does not look a central area that attracts and hence concentrates *inter* and *foreign* trucks although its delineation is quite clear for *intra* and *Belgian* trucks (Figures 6.11.c and 6.11.e).

The place of Brussels at a local scale

This Section aims to present the communities of movements observed in the former province of Brabant. Only the moves made from or to meshes located in the Metropolitan area are here considered. Like in the Section 6.4.2 (global scale), a benchmark is used in this Section (Figure 6.12.a). Figure 6.12.b shows the communities found in the "cleaned" general network. Communities detected are still made up of contiguous places, leading to a regional organisation. With the exception of the area located in the South of the

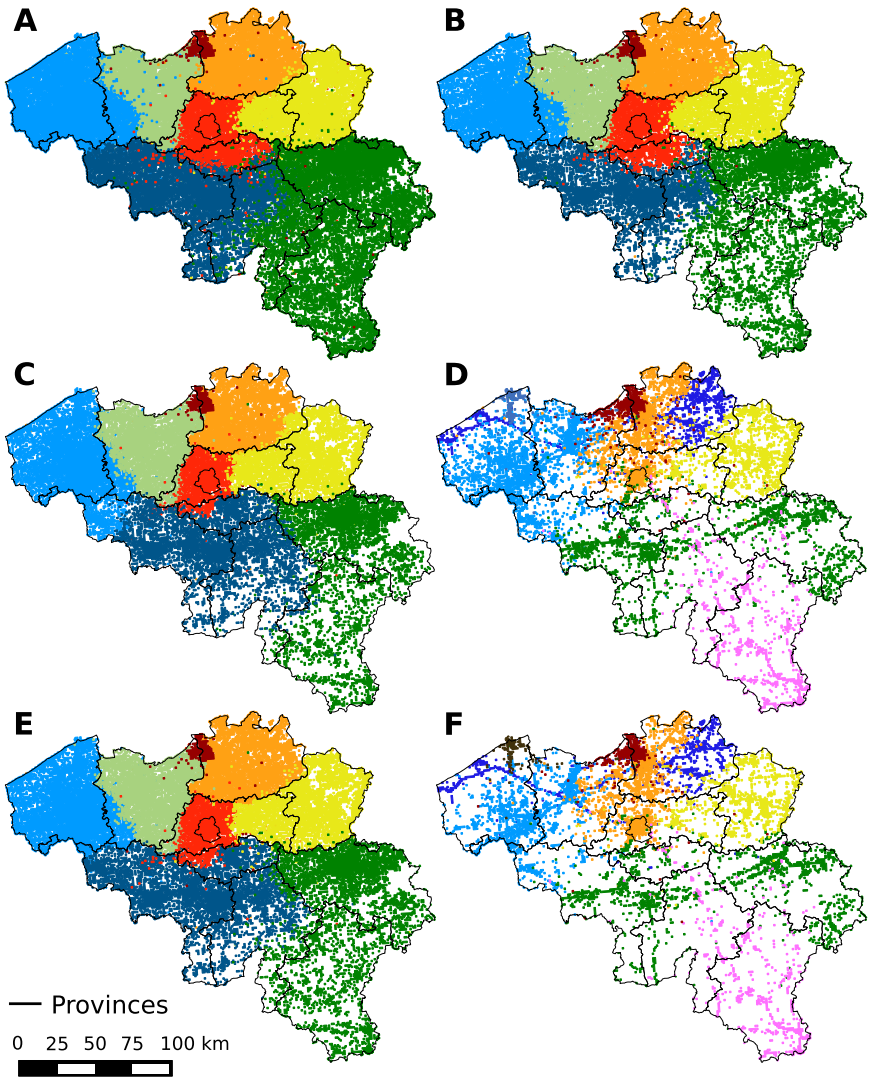


Figure 6.11. Communities of places located in Belgium and visited by truck: A) general network, B) places visited more than 10 times, C) *intra* trucks, D) *inter* trucks, E) *Belgian* trucks, and, F) *foreign* trucks.

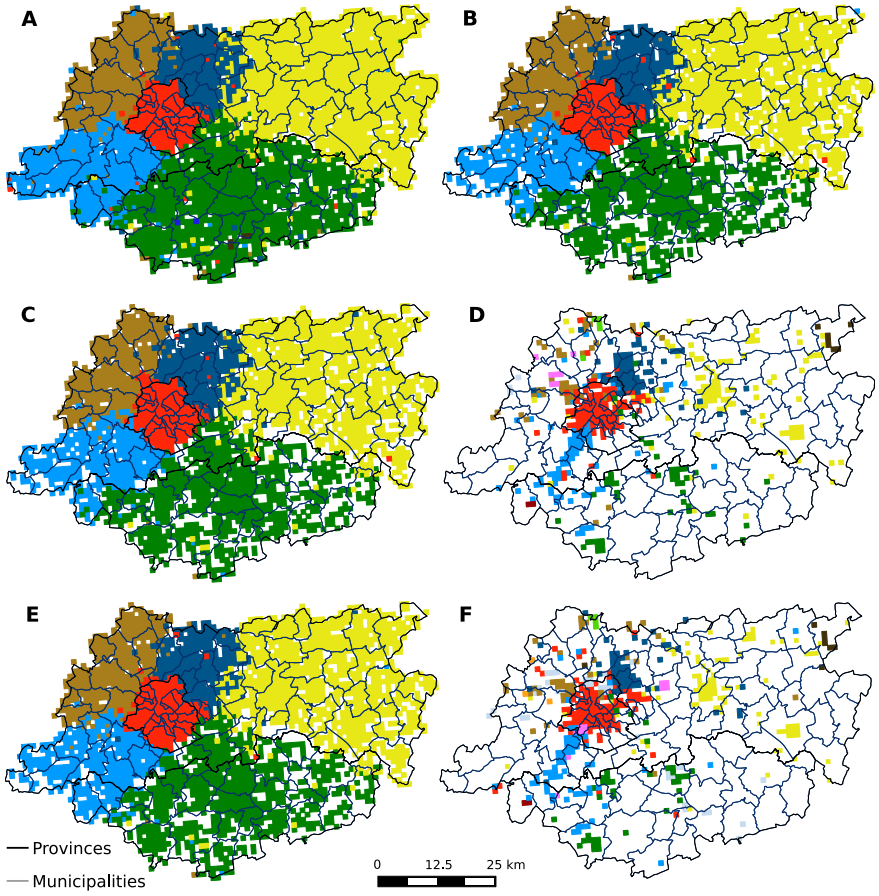


Figure 6.12. Communities of places located in Brabant and visited by truck: A) general network, B) places visited more than 10 times, C) *intra* trucks, D) *inter* trucks, E) *Belgian* trucks, and, F) *foreign* trucks.

BCR, the linguistic border is followed. Trucks have a higher propensity to make stops within only one province and do not cross the border between the North and the South. Similarly to the previous Section, *intra* (Figure 6.12.c) and *Belgian* (Figure 6.12.e) trucks have the same spatial partitions, expressing that although the networks are different, the realities observed within stops are similar. Communities detected in *inter* and *foreign* trucks are spatially well delineated: economic clusters, industrial zoning, city centres and the (air)port constitute the various communities. Brussels is mainly organised into three different communities located alongside the canal (South-West to North-East): the southern part along the canal, the central area within the BCR and the North organised around the port of Brussels and the airport of Zaventem. Other logistic centres, located close to Brussels, are found in the brown community (North-West side of Brussels) as well as close to Wavre where major companies can be found (dark green).

6.5 Conclusion

Transport geography has always dealt with a lack of spatial data centred on the moves made by the trucks. As regard of this contribution, GPS trackers implemented within trucks for taxation purposes reveals the usefulness of this data source for challenging transport geographical research. With small adaptations from their initial objectives, data accurately explain the spatial structures of the circulation of trucks at the scale of Belgium, as well as the economical polarisation of the Brussels area.

Brussels appears to be located at a central place in Belgium. This tendency to polarise a large part of the Belgian area is mainly explainable by the organisation of the circulation network organised around the capital. Moreover, Brussels and the Metropolitan area concentrate many economic activities where trucks stop for (un)loading purposes and hence correspond to a big economic market for companies which need trucks to transport their goods. Let us be clear within our explanations. Although this contribution uses the circulation of trucks as proxy to the circulation of goods, no information about the eventual load is available (neither in quantity or in nature - only the maximum loading capacity is given within the initial database). In addition, the nature of the stop taking place in a particular area is not considered in this contribution (unloading area, legal rest, comfort stops, etc). The hypothesis that analysing the circulation of trucks allows to follow the geographical component of the circulation of goods is not absolutely true (Pluvinet *et al.*, 2012).

Similarly, analyses made with the use of community detection method applied on different networks leads to different realities of the space. Belgian trucks (registered in Belgium or emitting GPS points only in Belgium) show

the regional organisation of the logistic sector in the country, although stops made by *foreign* trucks are more concentrated in specific places corresponding to economic clusters, urban or (un)loading areas. In addition, this contribution shows that, on the one hand, trucks registered in Belgium are not often leaving the country and are mainly deserving all the Belgian places raising question linked with the cabotage. Analysing the communities of places visited by truck shows insight for the case of Brussels: only considering international trucks do not lead to the detection of a community centred on Brussels.

To conclude, it is important to point out that many heavy cleaning steps were requested to ensure the reliability of the data. These operations ask an accurate question (Quesnot, 2016): do the choices made during the cleaning process considerably impact the data and hence the results? The determination of the stops based on a simple arbitrary time threshold is questionable. It would be more accurate to use spatio-temporal criterion taking into account the function of the places visited. Doing that would allow the exclusion of some stops (rest, comfort) based on the idea that they do not constitute (un)loading places of goods. Future research should be dedicated to detect these stops at the individual scale, thanks to the potentialities of new machine learning techniques that should be helpful to accurately understand individual spatial trace.

V

Movements of information

7

Brussels and its periphery: reviving the Chicago School with mobile phone calls?

Mobile phones operators nowadays provide large amounts of data, commonly used as proxy of social interactions, activities and mobility of the population. In addition to Parts II and III that focus respectively on people and freight movements, this Chapter analyses a third type of movements: information.

By delineating communities of antennas that are tightly connected in terms of calls at different periods of the time (Section 7.3.2), the objective is here to evaluate if groups of antennas are strongly embedded in space and if the temporal dimension affects the spatial organization of calls. As already expressed by Rodrigue *et al.* (2016), in addition of interactions, transport geography also requires the characteristics of the places (origins and destinations of the fluxes); this Chapter not only clusters antennas with comparable hourly variation of calls (Section 7.3.2), but also places with similar socio-economic activities (Section 7.4).

Finally, a concluding partition of the Metropolitan area of Brussels is proposed, based on interactions and spatial characteristics (Section 7.5); it gives a clear view on the three major aspects of the urban environment in Brussels: interactions, urban activities and composition of the places.

The research presented in this chapter has been carried out in collaboration with Prof. Isabelle Thomas and Dr. Olivier Finance and will be submitted to Environment and Planing B.

7.1 Introduction

Since the last 20 years, mobile phones are important items in our daily live. The purpose of mobile phone was to provide to the user, the ability to make/receive calls whatever the location nor the time. This initial objective is still present nowadays but new rising technologies were progressively installed within phones (internet, cameras, sensors are only few example), leading to what is currently called smartphones. By diverting specific sensors from their initial purposes, it is possible to access to data we could not imagine years ago: real-time positioning, real time sensors, social networks, and communication between people.

In recent years massive mobile phone location data have been studied, showing their great potential to model human mobility (Deville *et al.*, 2014; González *et al.*, 2008; Song *et al.*, 2010), to better plan urban areas as well as studying transport (Trasarti *et al.*, 2015; Calabrese *et al.*, 2011; Reades *et al.*, 2007; Järv *et al.*, 2018), to understand mobility patterns (Puura *et al.*, 2018; Sevtsuk and Ratti, 2010), to delineate phone basins as proxy of social basins (Blondel *et al.*, 2010), to explore the convergence with socio-economics data in an urban problematic (Cottineau and Vanhoof, 2019), or to improve statistical censuses (Debusschere *et al.*, 2017; Sakarovitch *et al.*, 2018). This rising interest of studying mobile phone may be explained by the difficulty to conduct classical travel surveys using self-report records (diaries or questionnaires), the lack of data about interrelationships (Duranton, 2015) or the accumulation of restricted assumptions (Dujardin *et al.*, 2007). These issues hence lead to the development of automatic mobile phone data collection systems (Trasarti *et al.*, 2015).

The temporal dimension is quite important when using mobile phone data. As short-term, specific events are easily detected and could hence be better controlled (Pucci *et al.*, 2015; Furletti *et al.*, 2017; Gabrielli *et al.*, 2015). On the long view, Trasarti *et al.* (2015) extract connections observed between places from highly correlated temporal variations of local population densities, while Calabrese *et al.* (2011) studied the general mobility of the population living/visiting the city of Rome.

As new rising technology, mobile phones allow to track users in a territory at fine spatial and time resolution by means of a telecommunication network. This kind of data provides a deeper information on the mobility of the population in shorter times and space (Gariazzo *et al.*, 2019), and although some bashes in spatial analyses (Kitchin, 2014; Rae and Singleton, 2015), interactions between places or agents (firms or households) can now be collected at the individual level (see e.g. Lee and Kang, 2015), allowing to accurately improve the understanding of the interactions of people within

urban areas, to the "interaction fields" proned by Fujita and Thisse (2013). Mobile phone data have turned out to be effective tools for analysis, visualization and urban planning (Hao *et al.*, 2015; Sagl *et al.*, 2014).

This chapter aims at measuring 3 aspects of the urban complexity: interactions, activities, and characteristics of the places within one urban environment: Brussels. We particularly aim at partitioning the urban space into sub areas (1) where phone interactions are more intense (community detection), (2) where places share the same characteristics (hierarchical cluster analyses) and (3) at further comparing and merging those partitions by means of a categorical cluster analysis. In addition, this Chapter joins interactions and characteristics of the places to determine what are the socio-economics components of each communities of antennas. Do people with the same socio-economics conditions have higher propensity to call similar people and do they interact with the rest of Brussels?

7.2 Studied area: Brussels and its surroundings

In order to avoid any discussion about the exact delineation of the Metropolitan area of Brussels, this Chapter focusses on the Former province of Brabant that is to say the BCR and the surrounding northern (Flemish Brabant) and southern provinces (Walloon Brabant). More than just avoiding hard discussion about delineation, this choice allows to include most of the Brussels urban core (Thomas *et al.*, 2012) as well as some other secondary cities and economical concentrations, such as Leuven, Vilvoorde-Mechelen, Wavre-Ottignies, Nivelles and Halle (these places are presented on each maps of this contribution).

7.3 Clustering places with mobile phone data

Let us now analyse mobile phone calls in two different ways: the extraction of communities of places with tight interactions, that are called *phone basins* (Section 7.3.2), and the delineation of clusters of places that look alike in terms of the time of calls, revealing groups of places with similar hourly variations of human (calling) activities within Brabant (Section 7.3.3).

7.3.1 Mobile phone data

Social interactions can take many forms and are not that easy to grasp statistically and quantitatively. ICT data have opened a new avenue for measuring and further understanding social interactions within territories.

We here use 13 million of mobile calls provided by one of the major operators in Belgium. The data set contains all calls made between antennas located within the former province of Brabant during 42 days in April and May 2015.

Due to privacy reasons, our data set is very simple: the number of calls made between each pair of antennas and the time of the calls. Calls are geocoded by antenna and no personal information about the customers or their call is provided. It is hence impossible to know if the calls were made for private or professional reasons, nor their content or their length. Also, the penetration rate of the phone operator is confidential and it is also totally impossible to compare our data with those of another operator. However, given the importance of the operator providing these data, we assume that there is no reason to expect major spatial biases at the level of Brabant. It is however important to insist: contrarily to census data, the results presented here do not represent the totality of the calls made, nor the entire population. Moreover, our data set only pertains to calls made between antennas located in Brabant, ignoring the "rest of the world". Such biases are inherent to most big-data sets (Longley *et al.*, 2015). Thanks to this new type of data we are able to discover new aspects of the urban complexity but we have to keep in mind that sensor data are here diverted for scientific purposes; they are not collected initially for that.

Dealing with mobile phone data allows to explore a temporal dimension (Sevtsuk and Ratti, 2010) and hence the variation in time and space of the human activity (Carter, 1981). The daily variation of the number of calls (Figure 7.1.A) clearly illustrates the differences between work day and week-end/holidays such as in our case Easter school holidays or special public holidays (1st of May; Ascension Day). There is also a clear daily variation during the week: the number of calls grows from Monday to Friday with a small reduction on Wednesday (no school in the afternoon in Belgium, leading to a larger share of part time working of the parents); a maximum is observed on Fridays (rush for the end of the working week and calls for organizing week-end activities). When looking at the hourly number of calls during a representative working day (21.05.2015), the variation is once again not random: almost no call at night, then the number of calls increases until 06/07 am with a peak at 12am, a reduction during lunch time as already demonstrated by Ahas *et al.* (2015) and Järv *et al.* (2014). The number of calls increases in the afternoon with the most important peak at 5 pm. After the working period, the number of calls decreases very rapidly to a value close to zero after 10:00pm. (Figure 7.1.B). Knowing the purpose of the call (private, professional) would have helped to better describe and understand this curve, but this is not possible nowadays due to privacy issues.

The location of the calls is mapped on Figure 7.2. Due to privacy reasons, several choices had to be made: the number of calls is here computed on a

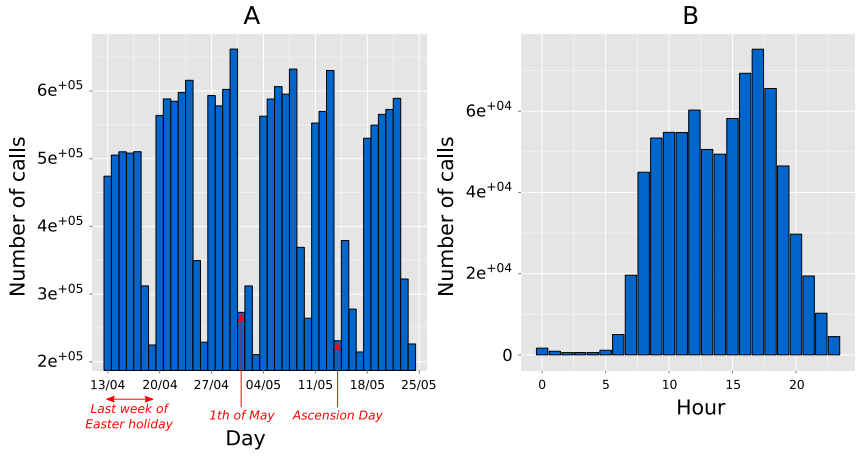


Figure 7.1. Number of mobile phone calls per day (A) and per hour on one working day (B).

grid of 3x3 km sized cells and each cell is colored in function of the percentage represented by the calls emitted/received by antennas located within the given cell in the total number of calls observed in the studied area. The discretization method used here is the equal frequency method; this method is easy to understand and is adequate for comparing magnitudes (Kraak and Ormeling, 2011). Large numbers of calls are located within and in the close vicinity of the BCR, including a south-west tail along the canal. Numerous calls are also observed in/around small Walloon regional centres such as Waterloo/Braine-l'Alleud, Wavre/Ottignies-Louvain-la-Neuve or Nivelles. Grid cells located in the North-Eastern part of the studied area have a lower number of calls. This spatial structure is explainable by the data themselves: beside the problem of the unknown penetration rate of the operator, we here only consider calls made within the former province of Brabant and ignore the calls made with places located out of the province. This automatically leads to an underestimation of the number of calls made by antenna located close to the borders (Andresen, 2010) or linked with other urban poles (this is particularly the case of Leuven, which was clearly mapped by Blondel *et al.* in 2010 (Figure 4.1): authors used data from another phone operator and showed that Leuven is closely related to Mechelen quite often and Mechelen is located out of the province and hence not taken here into consideration).

If grid cells help to describe the data, it is not possible to use them for community detection as several antennas can be located in one cell. As we are not allowed to map antennas individually for privacy reasons, Thiessen polygons were drawn around each of the 1,182 antenna and considered as the

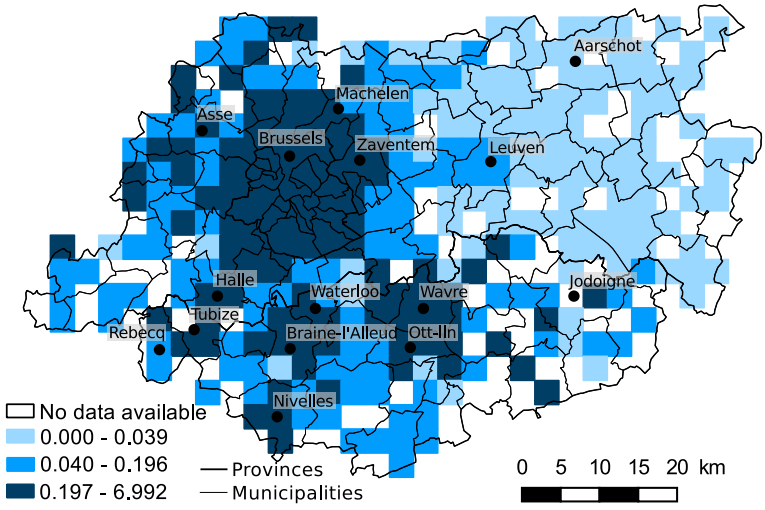


Figure 7.2. Location of the studied calls within the Former province of Brabant (in % of the total).

theoretical surface covered by the antenna. These polygons are for sure only a rough approximation of the reality as (1) they do not consider the real local physical or technical conditions of the field (buildings that reduce the range and strength of the antennas, etc.) and (2) they never overlap while in the real world two neighbouring antennas may cover a common surface (Zandbergen, 2009).

7.3.2 Clustering places with tight interactions

Method

We here apply the Louvain Method (see Chapter 2 for details about the method) to detect communities in four different networks: the first contains all the calls with no consideration of the time (*General*), while the others make a distinction according the temporal dimension: *Day* contains only the calls made between 08:00 am and 6:59 pm during working days, *Night* contains the calls made between 07:00 pm and 7:59 am during working days and *Weekend* corresponds to the calls made between Friday midnight to Sunday midnight. In order to quantify the similarities between the four partitions in communities, a NMI (Normalized Mutual Information) value was computed between each pair of maps (Equation 2.6 in Chapter 2). This index takes the value of 0 when the two partitions are totally different, and 1 when

the two maps are perfectly identical (Strehl and Ghosh, 2002).

Community detection (phone basins)

The partitioning into communities of the mobile telephony flow matrix measured between pair of antennas is represented in Figure 7.3. Similarly to Section 2.2.2, the municipalities characterized by the highest values of weighted degree (based on a Jenks classification) are represented on the figures with black dots located at the centre of these places. These municipalities correspond to centres of the computed phone basins. Let us bear in mind that we only have data when at least or the antenna of origin of the call, or of the destination was located in the province of Brabant; this is why we have chosen not to map what takes place outside the province of Brabant. In order to make the chorochromatic maps comparable, Thiessen polygons have been drawn around the antennas and have been classified as communities (Figure 7.3.A); these data have also been aggregated on the scale of municipalities and have been partitioned into communities (Figure 7.3.B).

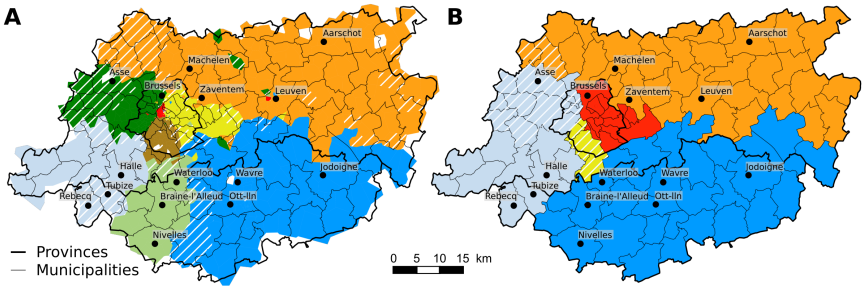


Figure 7.3. Mobile telephony areas on the scale of Thiessen polygons around antennas (A) and on the scale of municipalities (B).

The first striking feature is that even if mobile phones enable technically to call from anywhere to anywhere at the same price, communities of antennas within the province of Brabant are made of adjacent Thiessen polygons: calls are made more often between people who are near each other, confirming (Adam *et al.*, 2017; Blondel *et al.*, 2010; Expert *et al.*, 2011). Contrary to analyses made in Chapter 4, the spatial aggregation affects results slightly in this case. While the general appearance is identical, the optimal number of communities is slightly greater for the fine meshes (eight communities in Figure 7.3.A) than for the coarse meshes (five communities in Figure 7.3.B). Data aggregation erases spatial details in the case of telephone data: the CBD (Central Business District – heart of Brussels) no longer appears on the aggregated map (Figure 7.3.B), the community Braine-l'Alleud/Nivelles vanishes and the western part of the outskirts now forms a single community.

Aggregation implies a loss of information, which has an impact on spatial partition, while it did not have an impact on commutes, thus confirming other studies under way (Decuyper *et al.*, 2018).

A second feature when evaluating the Figure 7.3.A, is that Brussels is divided into pie-shaped wedges anchored in the city centre and extending to the periphery, crossing the BCR border, while the further periphery is divided into large blocks. In the centre, the CBD (in red) is very small but includes 4.5 % of the total number of calls and gathers a very large number of calls that are made in the offices located in multiple-storey buildings. The Brussels-Capital Region (outside the CBD) is divided into three communities (west, east and south) which extend spatially well beyond the perimeter of the Region in the same direction, sometimes with a little more uncertainty as the distance from Brussels increases (hatching). The explanation for this extension is related either to different socio-economic characteristics or to the fact that, by definition, mobile telephony is possible anywhere and does not involve a fixed workplace or a residence. Figure 7.3.A clearly reminds the sector model that Homer Hoyt largely described for Chicago in 1939 in terms of activities/land use and urban growth (Carter, 1981), a very traditional model in urban geography.

In terms of telephone relations, the regional boundaries (Brussels) and provincial/linguistic boundaries are not as well respected than they are in the case of commutes and migrations. For example, let us point out that the municipalities of Overijse, Huldenberg, Oud-Heverlee and the southern part of the municipalities of Bierbeek, Hoegaarden, Tienen and Landen are part of the telephone area of east Walloon Brabant: the linguistic boundary is therefore much more permeable in terms of telephony, than in terms of residential or pendular migrations. Tubize and Rebecq (French-speaking) are also connected to the telephone community of Halle (Dutch-speaking). Finally, let us note that in this case, the community of Leuven extends towards the municipalities in the north of the province up to Grimbergen.

In addition to the location, we also have information regarding the time of each call, thus allowing spatial and temporal analyses. The partitions into communities of the four networks defined above (see Section 7.3.2) are illustrated in Figure 7.4. The optimal number of communities detected within each sub-network is respectively 8, 6, 7 and 6. Chorochromatic maps represent the communities: there is no meaning of the color used but we tried to make the maps as comparable as possible through the subsets. Table 7.1 translates quantitatively the resemblance between the partitions by means of NMI values (see Equation 2.6 in Chapter 2).

Similarly to Figure 7.3.A, communities of antennas detected within different time dimensions are always composed of adjacent Thiessen polygons. People seems calling more closer friends/colleagues than distant ones. A

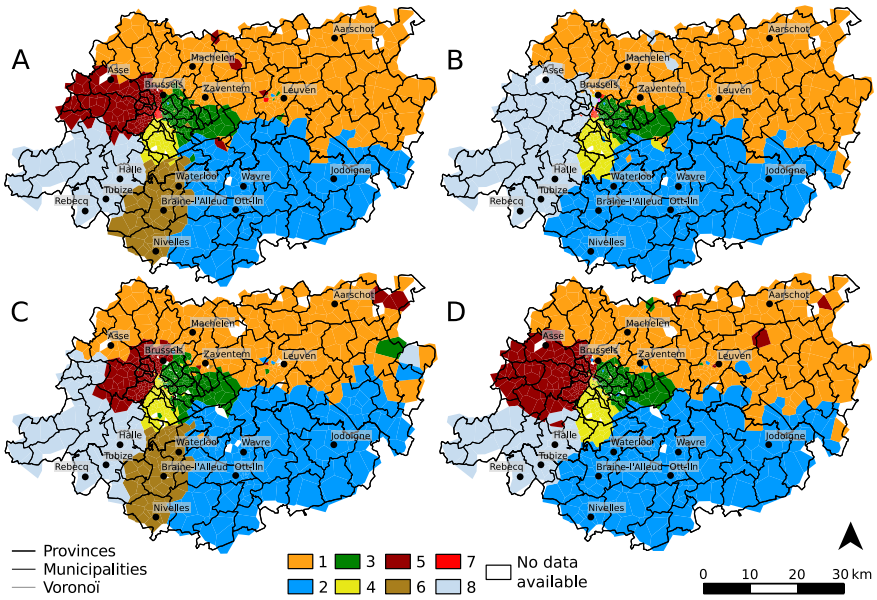


Figure 7.4. Mobile phone calls: communities of antennas detected for the four predefined networks: *General* (A), *Day* (B), *Weekend* (C) and *Night* (D).

second similarity is that the urban structure developed previously is observable at a main difference. Depending of the hour of the day, the Louvain Method does not detect a community in the CBD.

Table 7.1. NMI measured between pairs of calls network.

	Day	***Weekend***	***Night***
General	0.76	0.69	0.70
Day		0.58	0.75
Weekend			0.62

When comparing visually (Figure 7.4) and quantitatively (Table 7.1) the results obtained on the four networks, we see that all calls (*General*) and day calls (*Day*) look very much alike, simply because day calls are the most numerous in the total number of calls and hence drive the structure. All NMI values (Table 7.1) are larger than (or really close to) 0.6, meaning that the spatial structure of the interrelationships by mobile phone look alike, whatever the time period (4 networks). Week-end calls appear to be the most different from all others, while *Day* and *Night* have a quite large NMI. Major differences

between *Day* and *Night* are (1) the central community that disappears during night and week-ends confirming the office activity in this community, and (2) differences in the western part of Brussels where a large residential zone appears at night encompassing municipalities such as Dilbeek, Anderlecht or Molenbeek.

Figures 7.5.B and 7.5.D supplement this geography by illustrating the strength on partitions. The choropleth map is preferred here, as the Thiessen polygons are very small, making it difficult to see the hatching on the chorochromatic maps. The largest uncertainties are located in the west of the area under study but also affect some isolated places distributed randomly in the area under study.

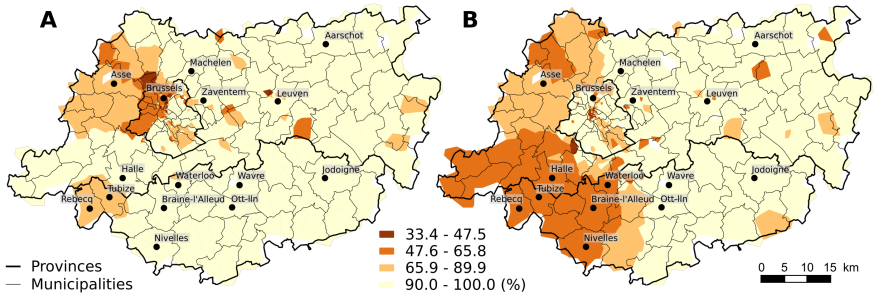


Figure 7.5. Communities of antennas: uncertainty of the classification of the day (A) and night (B) partitions.

Last but not least, the circular plot (Figure 7.6) confirms visually that people are mainly calling other places located in the same community: more of 50% of the calls are made within the same community at the exception of community 7 which is mainly linked with communities 3, 4 and 5 during the working period. Nevertheless, when the workers leave this area to come back to their place of residence, the number of calls decreases drastically, leading to the disappearance of this community in the *Night* and *Weekend* partitions. This graph confirms the veracity of the communities of antennas obtained by the Louvain Method: flows within communities are correctly maximized while the flows between communities is minimized (small fluxes in the middle of the graph).

Finally, Figure 7.7 proposes a close-up look at the central part of Figure 7.4.B (day) and another of Figure 7.4.D (night). For the sake of clarity, some place names are mentioned as reference points. Certain polygons form islands, classified as communities which are different from the ones which surround them, thus breaking with the first law of geography (Tobler) at this scale of analysis: each element interacts more with nearby elements than with far-off elements. For example, the antennas covering the areas of

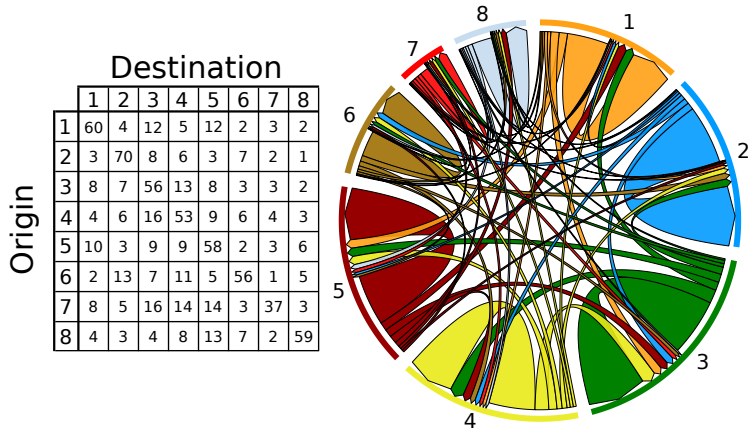


Figure 7.6. Circular graph and the associated O-D matrix (expressed in %) for all phone calls (*General*).

VUB and VRT are classified in the community of Leuven, while the antennas covering the Luxembourg, Schuman and Congrès stations are classified in the community of Walloon Brabant. These structures are detected for calls made during the day (Figure 7.4.B) as well as the night (Figure 7.4.D).

7.3.3 Clustering places with similar hourly variation

Pulses of the cities are nowadays better grasped using ICT data (Steenbruggen *et al.*, 2015; Trasarti *et al.*, 2015; Yuan and Raubal, 2012); busy city centres are very active during the day while almost dead at night; suburban residential areas wake up in the evening and during week-ends. Phone calls are an opportunity to measure these pulses. Figure 7.1.B gave a general shape of the hourly variation, while Figure 7.8.A represents the hourly number of calls for four given antennas: each antenna has its own temporal profile, some concentrate many calls in the morning while others communicate only during the working periods. In other words, each antenna has its own temporal pattern of calls emitted/received that will allow us to characterize places (Ahas *et al.*, 2015; Graells-Garrido *et al.*, 2016). This Section deals with the spatio-temporal dimension of these variations within the former province of Brabant.

In 1838, Verhulst developped the logistic function to model the population growth of natural species and will be further commonly used in biology (Gottschalk and Dunn, 2005; Herman *et al.*, 2008) or engineering (Villanueva and Feijóo, 2018). Some of its parameters are particularly relevant in our case (see Figure 7.16 for a graphical representation), such as the curve's maximum

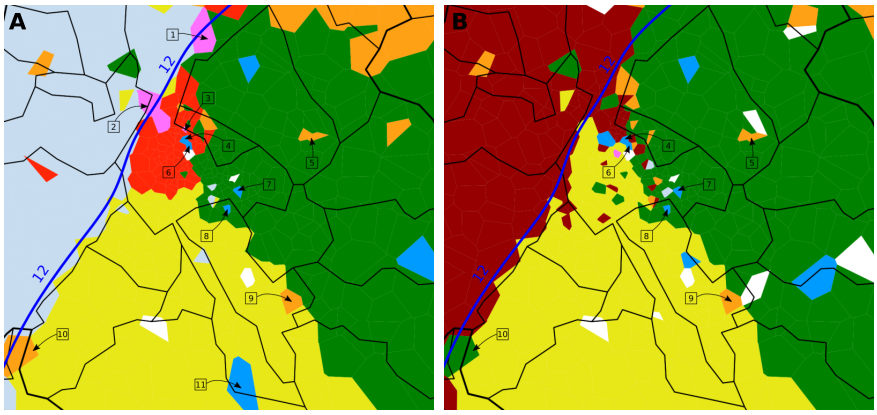


Figure 7.7. Mobile telephony areas (close-up of the centre of Brussels) during the day (A) and night (B) where locations on the map are: 1. Marché matinal et centre Européen de fruits et de légumes, 2. Quai des péniches; 3. Place Rogier, 4. Université Saint-Louis - Bruxelles (USL-B), 5. VRT - RTBF, 6. Gare Bruxelles Congrès, 7. Gare Schuman, 8. Gare Bruxelles-Luxembourg, 9. Vrije Universiteit Brussel (VUB), 10. Audi Forest and other industrial areas, 11. Bois de la Cambre.

value (d), the value of the sigmoid midpoint (g) and the steepness (b) of the curve (Bacaër, 2011):

$$f_x = 1 - \frac{d}{1 + e^{-b(e-e_0)}}. \tag{7.1}$$

To better isolate the problem, we here limit ourselves to working days, from Monday 00:00 to Friday 23:59; we end up with three complete weeks that do not contain any holiday. A logistic curve is then computed on the cumulated calls of each antenna and the value of the parameters b, d and g are extracted. It is clear that the majority of the antennas have the same profile: (1) few calls emitted during the night and early morning, (2) increasing number of calls from about 07:00 to 17:00, and (3) reduced number of calls in the evening.

Once the 3 variables d, g and b are collected, a Ward classification is applied in order to group the 1,182 antennas (Thiessen polygons) into clusters of antennas that have similar characteristics (Rokach and Maimon, 2005; Ward, 1963). The optimal number of clusters is found when the values of the Intra-Clusters Variance and the Entropy are minimal and when the Dunn and Calinski-Harabasz indices are maximal (Estivill-Castro, 2002; Ketchen and Shook, 1996). The values of these indicators were computed for a wide range of cluster numbers (from 1 to 30). The optimal number of clusters is here five and their spatial footprints are mapped in Figure 7.9. Clusters respectively include 515, 88, 280, 192 and 63 antennas. The average shapes of the curves are summarized in Figure 7.8.C.

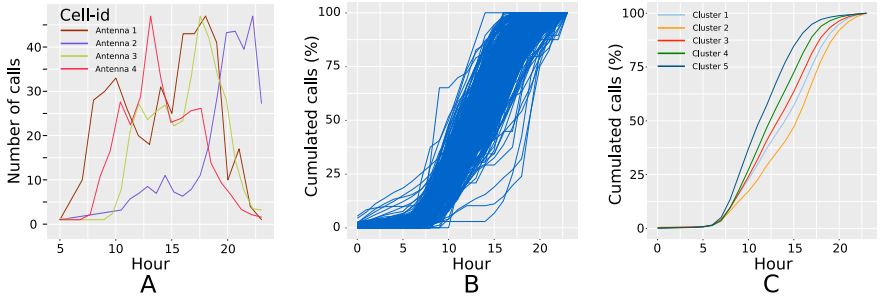


Figure 7.8. Temporal variation of phone activity: A: number of calls emitted by four randomly selected antenna, B: cumulated calls emitted by each antenna, for all antennas, C: average cumulated calls for each cluster of antennas.

The variations of the values taken by the parameters (Figure 7.10) and the general trends of the logistic curves (Figure 7.8.C) confirm that these five clusters can be associated with residential activities (Clusters 1 and 2), business activities (Clusters 4 and 5) and mixed activities (Cluster 3). Contrarily to the previous map (Figure 7.4), another vision of the living reality of the province is here provided.

Clusters 1 and 2 mainly characterize residential areas (De Maesschalck *et al.*, 2015) where most inhabitants are absent during working periods. Cluster 1 is characterized by a mean value of b of -0.35 and a sigmoid midpoint (g) located around 02:00pm. It explains that these antennas have a lower calling activity than antennas classified into the Clusters 4 and 5, as confirmed by Figure 7.8.C. Before 08:00am, the curve characterizing Cluster 1 follows quite well the tendency of the curves related to Cluster 3 and 4. Between 08:00am and 03:00pm, the number of calls made/received by each antenna classified in Cluster 1 is small. Finally, from about 3pm, people located close to these antennas start to communicate, which reduces the gap with the other curves around 10:00pm. Cluster 2 is quite similar to Cluster 1 at the exception that the value of parameter d is quite higher (after 03:00pm). These antenna emit/receipt a low volume of calls in the morning and the steepness of the curve drastically increases at 03:00pm; these areas are more active in terms of telephony after the working period and more especially during the evening.

Cluster 3 has intermediate values of the parameters. Indeed these antennas are located in quite densely populated areas where private housings stand alongside businesses (Dujardin *et al.*, 2004). Finally, Clusters 4 and 5 tend to characterize businesses. They both have a small value of g (around 11 and 12:30 am) explaining that these areas are mainly active in the morning and the early afternoon; afterwards, the amount of calls decreases. As a supplementary example, the curve of the antenna 4 (Figure 7.8.A) is characterized by a high

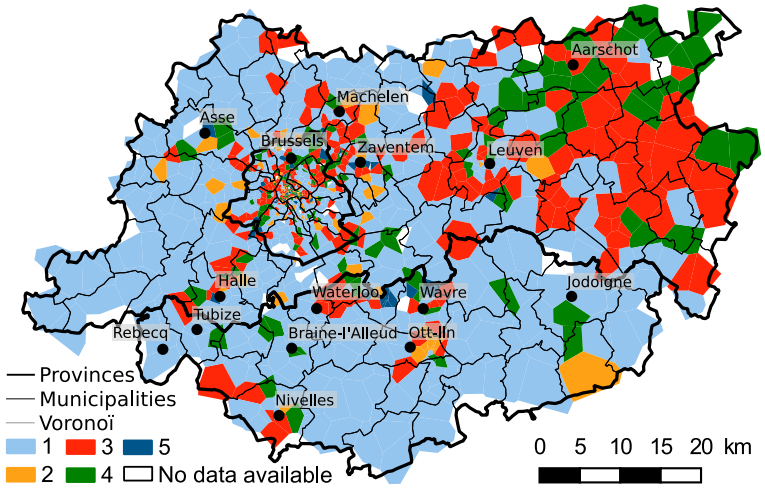


Figure 7.9. Clusters of antennas based on logistic curve characteristics of their hourly workday's activity.

volume of calls during the working period and is thus classified in Cluster 4. Cluster 5 mainly concentrates in (former) industrial areas, logistic centres or other economic activities with a strong spatial footprint, where the workers are especially active early in the morning.

Let us zoom on two areas in order to support the preceding comments. First, let us look in depth into the BCR (Figure 7.11.A). Residential areas are classified within Cluster 1 and at some parts in Cluster 2 for antennas more active during late evening such as locations of parties around the Ixelles Cimetery (point 1 on figure 7.11.A) and the Place Eugène Flagey, or festive activities within the historical city around the Grand Place of Brussels (point 3 on figure 7.11.A). The mixed types of antennas (Cluster 3) are mainly organized around the Canal in the centre of the BCR as well as in the North-East part (mainly residential units close to services and small commercial units). Cluster 4 is located around the Canal but at the extreme area of the BCR. These places are occupied by industrial activities with large footprints (e.g. car factory VW-Audi in the South and the Port of Brussels in the North - Strale, 2017). It is hence not surprising to see these places classified within Cluster 4 and 5 and located close to Cluster 3 to benefit from the positive externalities of these join locations. It is interesting to see that the areas with economic activities (Clusters 3, 4 and 5) are crossing the border of the BCR in the North by following the canal and spread in the Flemish area, thanks to the Brussels airport in Zaventem and the proximity of the Port of Brussels (Dujardin *et al.*,

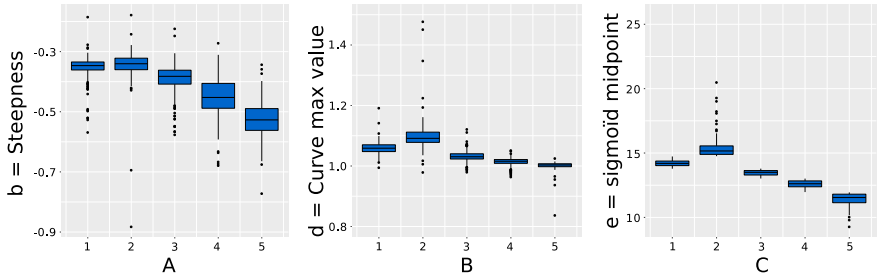


Figure 7.10. Composition of the five clusters of temporal curves: A: parameter b (steepness of the curve), B: parameter d (curve max value), C: parameter g (Sigmoid midpoint).

2004).

The second zoom (Figure 7.11.B) deals with the municipality of Ottignies-Louvain-la-Neuve which encompasses two realities: the western part corresponds to Ottignies (light blue) which constitutes the original municipal centre with a wide variety of intertwined commercial, residential and service functions, although the eastern part is a new town (Louvain-la-Neuve) developed since the 70s around the university. The western part belongs to Cluster 1 (more residential – light blue) and the eastern part is characterized by a mosaic of small Thiessen polygons included respectively into Clusters 2, 3 or 4. Louvain-la-Neuve is mainly characterized by various activity schedules associated to teaching activities, research centres, family residential areas (10 on Figure 7.11.B), student residences and party places. Antennas classified within Cluster 2 (orange) are located in the city centre where students attend courses and go to party at night. The main student residences area are located in the North side of Louvain-la-Neuve (11 on Figure 7.11.B). At the East side, antennas classified in Cluster 3 (red) are located further away and correspond to small student residences area with sometimes locations of businesses. Finally, the major industrial zoning area is classified in Cluster 4 (green).

7.4 Clustering places with similar characteristics

Let us make a very standard exercise in urban geography: depicting the structure of a territory by clustering places that look alike in terms of a set of selected characteristics X_i, reminding the so called "factorial ecology", a well-known technique anchored in early quantitative geography and urban modelling (Janson, 1980; Lebowitz, 1977). The objective is here to describe the province of Brabant and to delineate clusters of places that look alike and

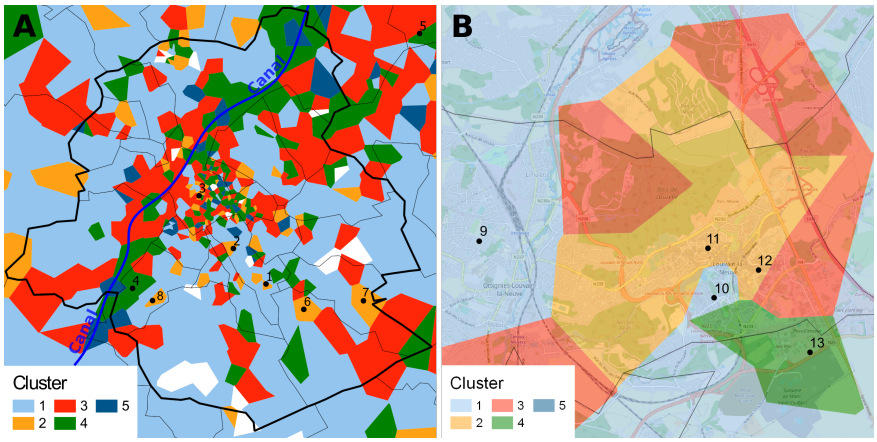


Figure 7.11. Zoom on Brussels (A) and Ottignies Louvain-la-Neuve (B) where locations on the map are: 1 the Ixelles cemetery, 2 is the Eugène Flagey Place, 3 the Grand Place, 4 VW-Audi car factory (Forest), 5 Brussels Airport (Zaventem), 6 and 7 Sport centres, 8 Concert Hall (Forest National), 9 Ottignies, 10 family residential area, 11 student residential area, 12 Place des Sciences (party place) and 13 Industrial zoning areas (Parc Einstein, Parc Fleming and Axis Parc).

more specifically to see if Brussels and its suburbs emerge in the constellation of smaller regional centres and periurban realities (Riguelle *et al.*, 2007; Servais *et al.*, 2004).

The characteristics of the places are provided by the latest census (Census-2011, 2016) at the level of the statistical sections (Statbel, 2019). Let us remind that each municipality is statistically divided into statistical sections and further sectors. Sections hence correspond to the aggregation of statistical sectors that are the smallest level of aggregation for which data are officially available in Belgium. Statistical sections were preferred to sectors because of their larger size, avoiding too many small/missing numbers. The studied area counts 982 sections among which 11 have missing values for at least one variable or count no inhabitant; we end up with 971 statistical sections.

A selection of six variables is used to characterize the socio-economic realities and the urban environment (see Table 7.2). Their choice is largely constrained by their availability at the section level and well inspired by theory. These variables are: (1) population density (*dpop*), (2) the percentage of youngsters ($<$ 15y) in the total population (*lessthan15 y.*), (3) the unemployment rate (*Unemploym*), (4) the percentage of commuters working in a section but residing in another municipality than that of the section (*commut.in*), (5) the proportion of dwellings built before 1946 (*builtbef.1946*) and, finally, (6) the percentage of dwellings occupied by a renter

(*rentedresid*). Unfortunately, it was impossible to use income data due to the too many missing values observed (small number of inhabitants) and it was decided to apply our analyses on specific proxy variables, like the unemployment rate and the percentage of dwellings occupied by a tenant (Timms, 1970). *dpop*, *commut.in* and *builtbef.1946* describe the urban environment where people live as well as the economic activities, while the other variables are used as indicators of the social conditions of the households (Janson, 1980; Sapena *et al.*, 2016; Timms, 1970). Variables are standardized and Table 7.3 shows that the variables are positivity correlated at the exception of *commut.in* that are negative (or more weakly) with the proportion of youngsters and that of old housings.

A Ward classification is applied in order to group the 971 sections into a small number of clusters based on their resemblance on the six selected variables (Rokach and Maimon, 2005; Ward, 1963). Following the method used in Section 7.3 (small values for Intra-Clusters Variances and the Entropy while the Dunn and Calinski-Harabasz is maximal), the optimal number of clusters is here 4. Figure 7.12, as well as Figure 7.15, summarize the composition and the delineation of the four clusters. Clusters 1 and 2 reflect quite well the urban structure within the former province of Brabant, where Brussels sprawls out of the limits of its administrative border (BCR). Cluster 1 (red on Figure 7.12) is the smallest cluster (98 sections) that mainly corresponds to the historical urban centre of Brussels (De Keersmaecker *et al.*, 2003; Eggerickx *et al.*, 2010). It is densely built, and characterized by a high proportion of old housings, mostly rented, and often quite deprived (Van Criekingen, 2006). Only three out of these 98 statistical sections are not located in the BCR. Cluster 2 (237 sections, orange in Figure 7.12) has a different population profile: less densely built, higher percentage of jobs occupied by workers coming from further away. People are commuting from other municipalities to join their place of work located within and around the major cities as well as within economic zoning, mainly located in the suburbs (Ermans *et al.*, 2018; Verhetsel *et al.*, 2018). Regional centres emerge (Leuven, Nivelles, Wavre, Ottignies) within less urbanized places in the North than in the South, but these orange sections also correspond to former municipal centres.

Clusters 3 and 4 are the two largest clusters and correspond to places located further away from Brussels and hence less urbanized: Cluster 3 (263 sections) is largely represented in the southern part of the province, while Cluster 4 (373 sections) dominates the Northern urban landscape. Both clusters correspond to places with a lower population density and a more periurban landscape (Thomas *et al.*, 2007). The main difference between Clusters 3 and 4 is that the dwellings are more recent in Cluster 4 as could be expected from former housing analyses and by different land use policy between Flandres and Wallonia (Vanneste *et al.*, 2008).

We obtain a very clear centre periphery structure (Cluster 1 and partially Cluster 2) around Brussels, and a periphery divided into a North/South organization: the Flemish periurbanisation (Cluster 4) and the Walloon one (Cluster 3). However some green coloured sections are found in the North and orange in the South, The Flanders/Wallonia difference is well marked but not clear-cut. Finally, the regional city of Leuven only slightly emerges.

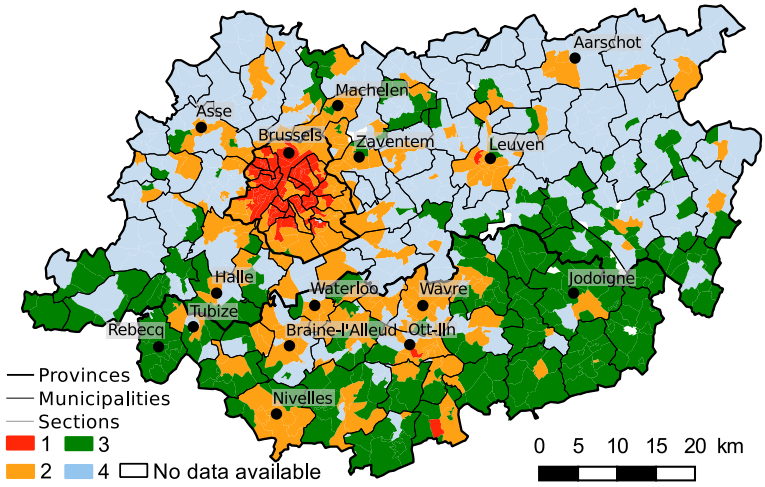


Figure 7.12. Clusters of statistical sections based on socio-economic variables.

If communities detected with phone calls (see Section 7.3.2) revived the Hoyt sector model, socio-economic basins and more anchored in the Burgess model (centre periphery, modified by political realities) and hence reminding the complexity of a urban territory where several spatial structures are layered.

7.5 Combining partitions

The objective is here to combine the above mentioned three partitions into one synthetic map (Sections 7.3.2, 7.3.3 and 7.4). Before looking at the results (Section 7.5.2), two methodological issues have to be raised and solved: the combination of the different basic spatial units used (Thiessen polygons and statistical sections) and the choice of a clustering method adapted for numerical variables (Section 7.5.1).

7.5.1 Methodological issues

Combining previous results into a simple and clear map requires to deal with different BSUs: Thiessen polygons (see communities of antennas and clusters of antennas with similar hourly variation) and statistical sections (see clusters of socio-economic condition). Combining these two different BSUs hence requires to conceive a unique and common spatial unit. Unfortunately, aggregation always leads to issues linked with a loss of information, and there is no "magic trick" to correctly keep the initial information. In our case, instead of aggregating phone information into the statistical level (like presented in the work made by Cottineau and Vanhoof, 2019), we decide to create an finer level of units that correspond to the intersection between the statistical sections and the Thiessen polygons (Figure 7.17 illustrate the process to create this new spatial unit i''). This kind of aggregation was chosen only because our two BSUs are characterized by nominal value (classification into communities or clusters) and not by cardinal values, which obviously would raised questions about the spatial distribution of the observations within the BSU.

Each newly calculated BSU belongs to an antenna community phone basin (Section 7.3.2), to a cluster of similar temporal patterns of human activities (Section 7.3.3) and to a Socio-Economics Cluster (Section 7.4). The problem is here that each BSU is characterized by a nominal value and not a cardinal value. With the emergence of big-data and Machine Learning techniques, several clustering methods have been adapted for nominal or even mixed data types (Aggarwal, 2014). An agglomerative hierarchical clustering data mining algorithm (called AGNES for AGglomerative NESting) is here applied. The algorithm initially considers each BSU as one cluster and then merges them into greater ones. The second step is repeated iteratively until only one cluster is obtained, when all BSU are member of just one single big cluster. For measuring distances between the nominal observations, a dissimilarity matrix is computed by means of the Gower distance (Gower, 1971): variables of k categories are first converted into k binary columns and then the Dice coefficient is used for computing distances between categories. The Gower distance varies between 0 (identical) and 1 (dissimilar). Six clusters are optimally obtained; their composition is mapped in Figure 7.14 and further detailed in Figure 7.13. Beside the analysis of the shape of the dendogram, two main indicators are used to find the optimal number of clusters: the within cluster and the Dunn index. The within cluster corresponds to the percentage of variance explained as a function of the number of clusters; the optimal number is detected when adding another cluster does not change much the data modelling. The Dunn index tells if clusters are compact or not (hence a small variance between members of the cluster, and, where the means of the different clusters are sufficiently different). A high value of the Dunn index indicates an optimal clustering.

7.5.2 A concluding map

Before analysing the spatial structure observed within the former province of Brabant, the composition of the different clusters are described (Figure 7.13). Each cluster of the abscissa is explainable by the three main previous classifications (socio-economics, community of antennas and the clusters of calls activity). For each of the main classification, the proportion of places categorized within each sub-cluster is coloured into a gradient of blue, where the dark blue represents 100%. Using Cluster 1 (abscissa) as an example, 75% of the places are classified within the socio-economics Cluster 2 while only 25% of the places are within Cluster 1. Figure 7.13 leads to a clear understanding of how the clusters in abscissa are organized in terms of the three main classifications. In the following lines, we propose a simple summary describing each of the six main clusters.

First, Clusters 1 and 4 fit well with the urban structure, made out of dense residential features alongside economical activities. Differences between these two clusters are due to the community of antenna showing that the Cluster 4 is more concentrated in well delineated places. Second, Clusters 2 and 3 do not include many economic activities and respectively represent the high dense residential areas and the periurban areas (outskirts of cities and countryside). Third, Clusters 5 and 6 correspond to residential BSUs alongside some economic activities. The main difference is due to the organization of the calls between places: Cluster 5 does not interact with the rest of the study area.

The spatial partition resulting from the classification (Figure 7.14) is consistent with the geographical knowledge on this area. First of all a clear North – South division is detected in Figure 7.14. But if most places in the North belong to Cluster 6 (blue) and most places in the South to Cluster 3 (orange), there are some exceptions randomly located. These two clusters mainly correspond to residential areas (De Maesschalck *et al.*, 2015), as confirmed by the presence of the Clusters 1 and 2 (Figure 7.9 and 7.13). Second, this cut between North and South is not well followed in the Western part of the former province of Brabant. A group of contiguous places is classified in Cluster 5 (green). At the opposite of the other clusters, Cluster 5 is quite well delineate in space without being too much constraint by the linguistic border: two Walloon municipalities (Tubize and Rebecq) are hence clustered with other Flemish places. The configuration of Cluster 5 is mainly due to the organization of the calls within the former province of Brabant. Figure 3 and 12 show that these places are highly connected in terms of calls and only a small number of calls leave Cluster 5. Third, the BCR is composed of four different Clusters. Each cluster is correctly delineated and fits well with the geographical literature available on Brussels. Among others, the historical city centre corresponds to Cluster 4 (red). This group is also

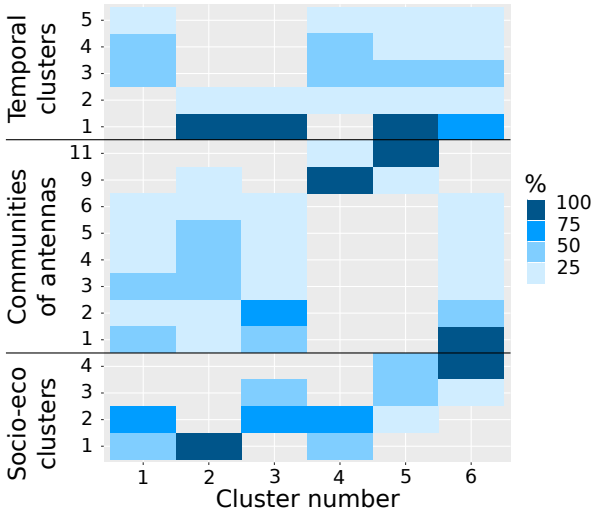


Figure 7.13. Composition of each cluster of the concluding partition mapped in Figure 7.14.

found along the canal around Halle and within the city of Leuven. Economical activities are clearly detected and classified within Cluster 1 (yellow): European corner, activities located close to the canal as well as within the Port of Brussels. The axis of the canal is a major industrial and economical axis extending to the North where companies are more interested in the space available instead of the concentration within the city centre (Dujardin *et al.*, 2004; Strale, 2017). Yellow places that are located close to the city centre contain many offices as well as head of private companies that choose the city centre of Brussels as a perfect location for their Belgian and European activities (Van Hamme *et al.*, 2016). This cluster is thus not limited to the BCR: it sprawls out of the border of the Region, from logistic zoning activity in the West side to zoning organized around the airport located in the East side of the BCR. This economical Cluster 1 is also detected in other places within the Former province of Brabant, such as around the cities of Leuven, Wavre or Nivelles. Close to the economic activities, is found the Cluster 2 (purple). It is exclusively detected within the BCR (except two units in the Walloon Brabant). It corresponds to purely residential places without economical activities, as illustrated in Figure 9 and 12. The places classified within this Cluster are characterized by a high density population, old residences and a high unemployment rate. In other words, this Cluster corresponds to places where peoples have socio-economic

issues (Van Criekingen, 2006; Van Hamme *et al.*, 2016).

The last cluster 3 mainly characterizes the Southern part of the BCR (as well as in some places located on the extreme North side). These units are mostly residential, made up of old houses built 1) with the extension of the city for places embedded in the Southern part of Brussels (Deboosere *et al.*, 2009), and 2) in the countryside located in the Walloon Brabant and hence corresponding to the former rural towns where the population density is small.

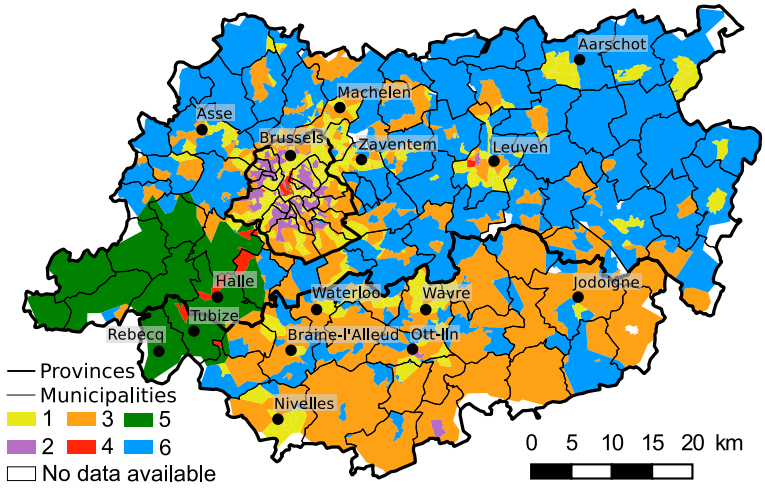


Figure 7.14. A concluding partition of the Metropolitan area of Brussels based on the three classifications.

7.6 Conclusion

We here confirm the usefulness but also the limits of unconventional data to understand urban space (spatial delineations and partitions). Mobile phone data are highly relevant to approximate quantitatively the relationships between people and to embed the time dimension within urban analyses. They complement advantageously conventional data and standard/traditional models.

Although mobile phones calls are not associated to wires, the cost of making or receiving a call may hence not be depended of the distance between two users. However the spatial imprint of the call locations is very strong: the composition of the communities of antenna are mainly made up of

contiguous places. In other words, close friends/colleagues call each other more often than distant ones. Moreover by definition, phoning is a time dependent activity and these variations reveal the pulses of the city; clustering antennas with similar hourly profile, enables one to show how residential and economical activities clearly partition the studied area. A multi-level structure emerges, consisting of several sectoral areas anchored in Brussels and reminding the so called (sectorial) Hoyt model. However, one has to keep in mind that phone data are not collected originally for measuring urban realities, and we hence do not know the length of the call nor their content. In addition, several phone operators share the market without any clear view on the penetration rates. Partitions provided by phone data were further compared to standard clustering of places by means of a selection of socio-economic characteristics of the places; it leads to another vision of the studied area, in phase with the well-known (radial) Burgess model in and around Brussels. The core and the outskirts of the city of Brussels are perfectly delineated, as well as the strong regional effect of the linguistic border (Thomas *et al.*, 2017) and the relative importance of regional cities such as Leuven, Wavre or Nivelles. An attempt to combine all the partitions (interactions, time variations, socio-economic characteristics) lead to an interesting spatial organization, confirming the internal complex organization of the urban agglomeration of Brussels (an historical city centre, an economical axis along the Canal and around the European Corner surrounded by residential places, etc.) and the difficulty for informations to easily move within this mosaic of multiple divert socio-economics realities.

More than just characterizing places, we are now more aware of how places, and hence peoples, interact through space and time. Combining flows (mobile phone calls) and characteristics (census datasets) allow to understand and summarize the characteristics of the places as well as the interaction of people within their environment (Roy and Thill, 2004): it brings new insight into the studying of the movements of information through space. By doing so, a step forward is made in the analyse of the urban complexity: instead of analysing separately different layers of the urban features, this geographical contribution reconciles the aspects that were not often analysed together.

Appendix

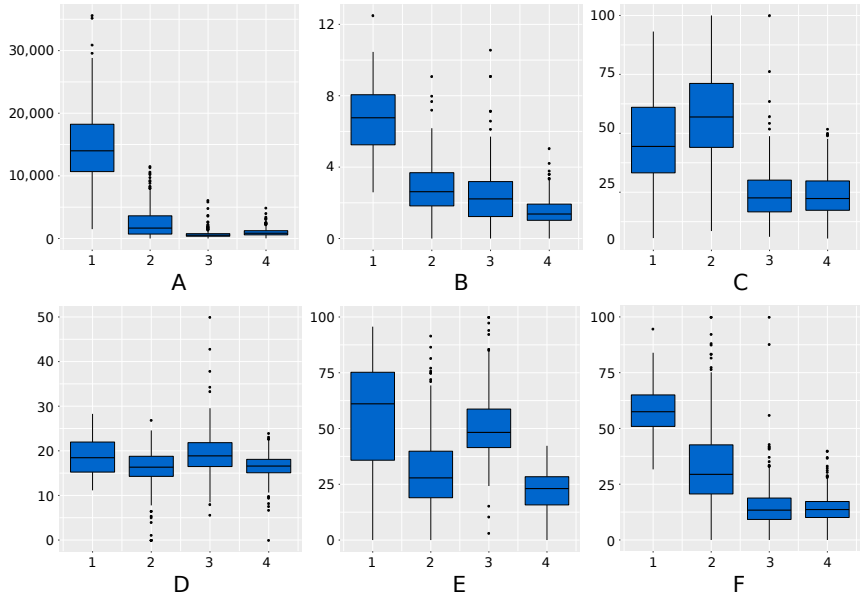


Figure 7.15. Cluster composition: A: dpop, B: Unemploym, C: commut.in, D: lessthan15 y., E: builtbef.1946, F: rentedresid.

Table 7.2. Definition of the variables.

Variable	Definition	Units	Source	Mean	Min	Max
Dpop	Population density	inhab/km^2	Census 11	2,690.53	2.22	35,620.08
Unempl.	Rate of the population registered as seeking a job within the active population	%	Census 11	2.58	0.00	12.5
Commut.in	jobs in a statistical section occupied by a worker not residing in the municipality encompassing the section.	%	Census 11	34.98	5.35	100.00
Lessthan15 y.	Population aged less than 15 years	%	Census 11	17.34	0	50
Built bef. 1946	Dwellings built before 1946	%	Census 11	35.12	0	100
Rentedresid	Rented dwellings	%	Census 11	23.89	0	100

Table 7.3. Pearson correlation coefficients between the variables.

	dpop	*unemploym*	*commut. in*	*lessthan15 y.*	*builtbef. 1946*
unemploym	0.709				
commut.in	0.264	0.267			
lessthan15 y.	0.106	0.188	-0.142		
builtbef.1946	0.305	0.415	0.101	0.251	
rentedresid	0.669	0.602	0.559	-0.014	0.258

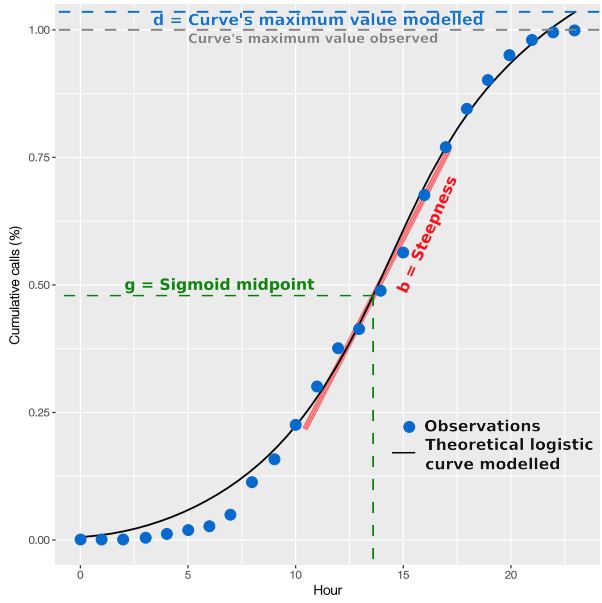


Figure 7.16. Theoretical logistic curve.

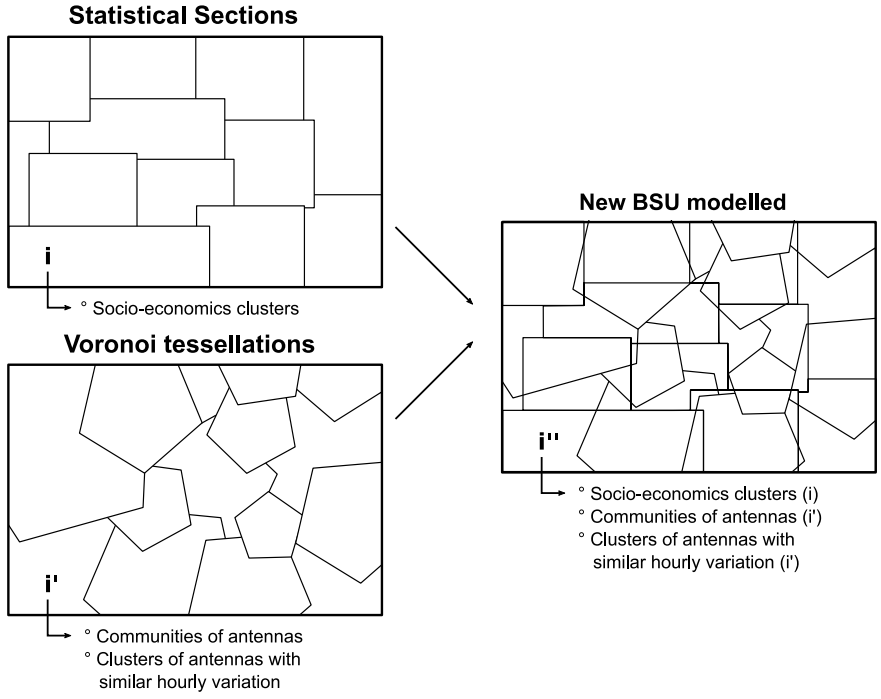


Figure 7.17. Creation of the new BSU based on two different spatial entities.

VI

Concluding remarks

8

Conclusion

This Chapter concludes this thesis by coming back on the general research question: to measure, understand and explain the interactions observed in and around the metropolitan area of Brussels by analysing networks of movements constructed from large (un)conventional OD matrices.

Section 8.1 presents the methodological choices made, as well as a general overview of the most striking results. In particular, it evaluates if the different groups of places are consistent with the current Belgian administrative organisation, or if huge spatial differences are observed. In addition, explanations were proposed to better understand how places and agents interact through space. Then, Section 8.2 proposes a critical review of the works made through this contribution, while Section 8.3 draw few paths for future research. The Section 8.4 provides a set of recommendations, useful for urban planning, territorial management, or decision making. Finally, the Section 8.5 draw the final conclusion of this doctoral research.

8.1 Summary of findings

8.1.1 General objectives of this thesis

This doctoral dissertation aimed to measure, understand and explain the interactions observed in and around the metropolitan area of Brussels by analysing networks of movements constructed from large (un)conventional OD matrices.

Studying OD provided by big-datasets, requires new techniques and community detection methods (Section 1.4), are useful to find how places are connected, and what their spatial extents are. To this end, and without any consideration of the characteristics of places, this thesis evaluated if these methods may provide useful spatial partition of the territory within movements of people, freight and information. Do they provide new insights into our traditional transport geography models?

8.1.2 Data and methodological framework

One of the facets of this thesis resides in the use of census and big-datasets to analyse interrelations between places: amongst others, commuting movements, mobile phone calls, train schedules requests or spatio-temporal tracking of trucks were collected and analysed.

A particular attention was given to the identification of the methods, on their use, as well as their improvements. The approach chosen to analyse interrelations appearing in Belgium and in the Metropolitan area of Brussels seeks to partition space into communities of places that are tightly connected. The most intense exchanges are hence located within communities instead of between them. Although this thesis is centred on the case of Brussels, analyses were firstly applied on Belgium, contributing to validate the place of the city in the Belgian urban structure, and avoiding the impact of the well-known "border effect".

The Louvain Method was particularly applied all along this contribution. Given that results are highly depended on the method, Chapter 2 presents the sensitivity of the Louvain Method to different geographical and methodological issues. An important aspect of this sensitivity analysis was the implementation of the resolution parameter ρ in the Louvain Method. This parameter modifies the size of the communities detected, from a large number of small groups to a small number of very large ones. This implementation highlights how the interrelations are embedded at different levels of resolution, and hence how interactions are organised in space. Like other classification methods, the way of grouping places depends on the objectives: classify a space into a determined number, or let this choice to the algorithm itself. In this contribution, we

make three major choices to select the value of the parameter ρ: (1) determine the optimal value(s) based on the linearized stability procedure developed by Delmotte *et al.* in 2011 (Section 2.3.1); (2) select a given number c of communities to make the result comparable with another partition already composed of c communities; or (3) using the standard Louvain Method which does not implement this parameter ρ, leading to an unknown, but optimal (considering the standard modularity), number of communities.

The level of aggregation is always a major geographical challenge because it induces a loss of information. For the purpose of being free of administrative units, this contribution used various levels of aggregation: statistical sectors, Thiessen polygons, municipalities, as well as a regular grid constituted by cells of $1km^2$.

In addition to the Louvain Method, several other methodologies were applied to delineate the interaction fields. The Infomap method was used in Chapter 2 to delineate groups of places tightly interconnected in terms of commuting movements, while a hierarchical clustering method, the Anabel Method, was presented in Chapter 3 in the problematic of residential moves in Belgium. Using different methodologies with different objectives, obviously leads to results that are distinct.

Another contribution of this thesis was to confront the computed communities with characteristics of the places studied, leading to the concept of transport geography. By applying "traditional" clustering methods, space was delineated into groups of places that have similar characteristics. Particularly, Chapter 7 groups together places having similar socio-economic conditions or having a similar temporal variation of calls. Finally, a concluding map combining interactions and characteristics of places was proposed at the scale of the Metropolitan area. It provided a clear vision of the spatial organisation of places in and around the BCR.

8.1.3 Key results of this thesis

In order to present a broad selection of the major results, two key questions will guide the development of this Section. By looking back on the analyses made within this doctoral research, do life, jobs, economic or communication groups detected, correctly correspond to the current administrative levels? Section 8.1.3 and 8.1.3 respectively discuss the administrative organisation of Belgium and Brussels with the objective to evaluate if administrative entities fit with the effective territoriality of life and economic areas.

At a national scale, interactions mainly occur within provinces

All along the analyses, and whatever the type of movements studied, results show that communities are mainly made of contiguous places, although no spatial information or characteristics of the places were integrated in the method. Close places interact more than distant places, confirming the first law of geography (Tobler, 1970). In addition, limits of these groups mainly correspond to regional/provincial limits and the linguistic border. These limits act like impediments to interactions: the difficulty to cross these borders results from a high cost (money, time or psychological), or from a bad knowledge of the other territorial entity (Nijkamp *et al.*, 1990; Grasland, 1999). The next paragraphs will particularly focus on the three kinds of movements (moves of people, freight and information) analysed, and assess if their spatial extent fit with the various Belgian administrative levels.

Firstly, communities detected in commuting and migration movements have different spatial footprints: population have a high propensity to move on longer distances to reach their places of work, at the opposite of residential moves. It leads to the detection of smaller communities while studying changes of residences. This may be explained by the positive externalities of their previous location: staying close to their families, to their "territories" and to their social networks (friends, leisure activities, schools, etc.). Although the size of the communities are different, the spatial organisation of commuting and residential groups followed quite well the urban hierarchy (Vanderstraeten and Van Hecke, 2019, , see Figure 4.5).

By collecting train schedules requests, it is important to notice that the OD matrix analysed corresponds to potential movements. Detecting communities of train stations shows that the computed groups are articulated around the main central cities. They attract travellers due to their large range of offers (work, leisure and tourism are only few examples). These communities have large spatial footprints that do not depend on the nature of the trip (work or leisure purposes that are approximated by time): partitions detected during the working days or the weekend show similarities, with the exception of stations around Antwerpen and Ghent. The contiguity property is still quite clear, but on the one hand, due to the organisation of the train network, the provincial limits are fuzzier and communities may extent in several provinces (the two Flanders, Liège-Namur-Brabant Wallon). On the other hand, linguistic border clearly affects the travel demand for train transportation: people from one linguistic Region do not know what happens on the other side of the linguistic border. A high number of stations from the other linguistic Region do not appear in the subset of a given linguistic group.

Secondly, this thesis analyses freight movements, approximated by the

circulation of trucks in Belgium. Applying the Louvain Method on these movements of trucks shows the clear effect of the provincial limits, as well as the linguistic border. Communities better follow the provincial limits in the Northern part of the country (with the exception of the area of Leuven), while Wallonia is mainly dominated by two communities (East and West) and where the Walloon Brabant is classified with the Former arrondissement of Brussels-Halle-Vilvoorde. This regional difference may be explained by a stronger density of inhabitants and a denser transport network in the northern area, leading to a better organisation of the freight sector into provinces (or couple of provinces).

By considering only movements made by truck registered abroad, the spatial footprints of communities are sometimes split on different parts of Belgium: a given community only found close to the borders of the Netherlands and France. This lack of contiguity is explained by trucks crossing Belgium during their travels with no consideration of the Belgian space.

Finally, although mobile phones considerably decrease the friction of the distance (the cost of a call does not depend on the distance between users), analyses reveal that social networks are still mainly ruled by distance and that people we are calling (friends, colleagues, companies, etc.) are located in a narrow radius, leading to communities made of contiguous places. Once again, reminding us of the first law of geography (Tobler, 1970). Like other analyses, and as expected, the linguistic border affects the communication pattern of people: it is obvious that callers have a higher propensity to call people from the same language community. Nevertheless, due to the proximity of Brussels and its mixed population (Dutch and French-speaking workers), the linguistic border is more permeable to calls that any other data sources used in this contribution.

At (intra)urban scale, limits of the Brussels Capital Region do not match with spatial behaviour

Analyses conducted on the interaction and characteristics of the places show that most of the results converge to an inadequacy between the limits of the BCR and the functioning of the city. The administrative border of the Region rarely fit with the limits of the computed communities. Whatever the kinds of interactions, communities sprawl in neighbouring areas located in Flemish and Walloon Brabant. But, if a more consequent number of communities is detected, the Region itself is organised into several small groups having well-distinct realities. Within the current debate of merging the 19 Brussels municipalities, and if an internal administrative organisation should subsist, our observations could bring insights to the discussion. Nevertheless

proposing a unique answer combining the various groups detected in and around Brussels is not trivial. Not only because the datasets represent different realities but also because each sample of the population is not similar: active population in the case of commuting movements and totally unknown when considering mobile phone data.

When analysing commuting movements, the community located in Brussels has a large spatial footprint that includes the Walloon and Flemish Brabant (with the exception of the area around Leuven), fitting well with the former arrondissement of Brussels-Halle-vilvorde. This large polarisation of Brussels may be explained by the inadequacy between the supply and demand for jobs within the city. Although this offer is quite high, the population of the BCR is characterised by the highest unemployment rate in Belgium (Dujardin *et al.*, 2008, , see Figure 7.12). Most of the workers are hence coming from Flanders and Wallonia to join their working places. These interactions between Brussels and its periphery connect places for economic reason. However, these movements are theoretical: we know that people are moving from their residence to their work place but there is a lack of accuracy. For instance, no information about part-time workers are provided.

Migration movements explain other stories: residential communities have small spatial footprints, leading to a delineation of the BCR into four major communities. The two largest ones are located on both sides of the canal, revealing a clear discontinuity within the Brussels urban fabric (Section 1.1). These two groups correspond to distinct socio-economic realities: the poorest population mainly located in the Northern part, and the richest population mainly found in the Southern part (Van Criekingen, 2002; Observatoire de la Santé et du Social Bruxelles, 2006). In addition, the principles of gentrification and pauperism force people to stay in places linked with their socio-economic condition (Van Criekingen, 2006; Van Hamme *et al.*, 2016). It ends up with only few moving between each side of the canal. On the East side, Evere, Schaerbeeck and Saint-Josse-ten-Noode are classified into a well-delineated group, while, the two "Woluwé" are classified within a community that sprawl out of the limit of the BCR, in the direction of Kraainem and Diegem.

This partition of the BCR may be linked with phone basins coumputed in Chapter 7. Spatial organisations are similar in both cases: a clear cut along the canal (poor and comfortable areas) and a third community of antennas extending out of the provincial limit toward Kraainen and Diegem. These observations confirmed the importance of mobile phones in our daily life and social networks (Deville *et al.*, 2014; González *et al.*, 2008; Song *et al.*, 2010; Blondel *et al.*, 2010).

Using another dataset, leads to another spatial reality. Detecting communities in requests of transport demand shows that train stations

located in the South of the BCR are more linked with Walloon stations while the North and West part is linked with the Flanders area. We consider these movements as potential because we have no clue that the request ends as a real use of train transportation. Nevertheless, for certain period of the day, we could hypothesise that these movements are made by commuters travelling between their places of residence and work. With a better control of why the request was made (checking the train table, controlling issues on the train network), it would be possible to detect structures consistent with ones detected in the commuting movements (Chapter 4) and hence presenting another point of view of this kinds of move.

Applying the Louvain Method on mobile phone calls reveals that communities of antennas have strong and well delineated spatial footprints that vary in time. During the day, the CBD is detected as a small community located in the historical city centre, and this spatial structure is vanished when studying calls made during the night or the weekend. It hence confirms the strong importance of the residential population in the communities of antennas detected, linked with the nature of the flow measured: communication. This spatial computed partition fit well with the sectoral description modelled by the Chicago School (Section 1.5).

Tracking trucks in real time was useful to understand how the freight movements between places are organised. Two major distinctions have to be made: (1) considering *Belgian* trucks leads to a community that closely follows the limit of the BCR, and (2) communities of places visited by *inter* trucks are mainly organised along the axis of the canal where most of the logistic activities is found (Macharis *et al.*, 2014). In addition, places alongside the canal are important locations for the road transportation. This axis is clearly followed and extents from the south-west to the north-east, explaining that economic activities are not constraints by the limit of the BCR. The major economic importance of these locations were already detected through the activity of antennas (Figure 7.14).

Considering characteristics of the places also leads to an inadequacy between the various delineations computed and the limit of the BCR. Firstly, a socio-economic partition of the Metropolitan area was made in Chapter 7. Using "traditional" clustering methods, space was delineated into groups of places that have similar socio-economic conditions. It leads to a concentric organisation of the BRC, remembering the well-know Burgess model (Section 1.5) and the historical development of Brussels into different crowns (Dujardin *et al.*, 2008; Van Criekingen, 2006; De Maesschalck *et al.*, 2015, and Section 1.1). The effect of this historical organisation is still strong nowadays, especially by constraining the migration movements (Chapter 4) and mobile phone calls (Chapter 7).

Secondly, by modelling the daily variation of the activity of antennas, space was clustered into groups of antennas that have the same daily pattern of calls. It leads to accurately determine the various typologies of residential or economic places, and hence to provide a better understanding of how the population really "live" in the city. It also highlights the economic places located alongside the canal, corresponding to logistic places (Macharis *et al.*, 2014), and hence confirmed by results obtained in Chapter 6.

Finally, following trucks in space brings new insight on the traffic (places crossed by truck) and the connections (places where trucks are making stops). The simple comparison of these two components reflects two distinct realities. On the one hand, the roads highly frequented by trucks are located on the outskirts of the city or along the canal, while, on the other hand, the majority of stops made by trucks are found all along the canal, from Halle to Vilvoorde, as confirmed by the community detection and the daily patterns of the antennas (Macharis *et al.*, 2014).

As a general observation, by the combination of interactions and characteristics, the limit of the BCR seems not well adapted to the various spatial structures observed through this thesis. These observations show that thinking and managing the BCR with no consideration of the close periphery is currently not adapted. Using limits that are not consistent leads to issues when considering mobility, urban planning, demographic transition and economic health of the Capital of Belgium. These challenges could be fully solved only when considering every interaction as a whole, and as an objective of better stick the way "users" really live the city.

8.2 Critical assessment

A doctoral research is a long project where many choices have to be made depending on the availability and the structure of the data. These choices sometimes affect results, and readers have to be aware of the various impacts. This Section proposes to have a quick critical assessment on the choices made through this contribution. Although limits of census data are well defined and known, data collected through other sources require caution. Issues are linked to the lack of exhaustiveness, wrong samples of the population, or choices made during the cleaning processes. Please read Longley *et al.* (2010) for a general review of problematic raised by spatial big-data.

The *iRail* dataset offers methodological challenges as well as an important potential for future research. For the former, two main limitations and critics should be outlined. First, a travel request made on the *iRail* website or application does not mean that the journey was actually made. The number

of requests cannot, therefore, be easily translated into forecasts of the absolute number of passengers. The second difficulty is that *iRail* is only one of the various schedule-finder websites and applications existing for Belgian railways. In addition, no information related to the characteristics of people are available, leading to a sample of the population which is not controlled. The analysis proposed in this contribution would gain to correctly know the limits of using a unique sample provided by only one data source.

Unfortunately, analyses based on mobile phone calls suffer from the same limitation. The dataset was provided by only one major Belgian company, and no information about the socio-economic conditions of customers were provided. Due to the specific subscription offers made by each phone company, it is likely that only a certain part of the population is registered (middle active class receiving the subscription from their work or only young people interested in internet connection). It is hence difficult to generalise our specific results to the Belgian population.

A second issue of this dataset is that mobile phone calls are aggregated by antennas, and only the total number of calls made between pairs of antennas per hour are provided. In order to make readable maps, it was decided to use a surface variable (Thiessen polygons) instead of a punctual one (location of the antennas). This theoretical coverage was the best solution in our hands to manage this location issue. In addition, it is well known that the real area covered by a given antenna depends on the physical features such as buildings, the landscape or the population density. Moreover, in the current mobile phone infrastructure, antennas have areas that overlap between them. Doing so, users are not always connected to the closest antenna, and events occurring in a particular location are wrongly matched to a neighbouring phone receptor. The wrong matching to the closest antenna and the overlapping property of the areas may raise a major issue. It is likely that if a user is located within an overlapping area, calls made by this user may be randomly distributed to various antennas although his real location never changed during the day. This effect would probably strengthen the strong spatial contiguity observed within communities. Using the real coverage of antennas may resolve these allocation issues.

Finally, no calls linked with places located out of the Former province of Brabant were available, making our partitions quite dependant of the border effect. We here only measure internal calls and ignoring phenomena appearing out of our delineation.

Analysing GPS traces made by the circulation of trucks is a striking challenge. Although this dataset was provided by a federal institute, it contains several kinds of errors that are listed within the Chapter 6. Three

main critics could be raised here.

First, the circulation of trucks is used as a proxy of the freight transportation. This hypothesis is not relevant because no information was provided about the real transport of goods. It is hence impossible to know if a given truck is carrying a load, and hence if it stops at a specific place for (un)loading purposes. This issue should be solved by incorporating characteristics of places such as land use information. In addition, only the movements made by truck of more than 3.5 tons were analysed. We hence have no insights related to the circulation of vans, yet important in urban context.

Second, using a single time threshold of 10 minutes to determine a stop is highly questionable. This threshold is chosen to be significantly low to catch different kinds of stops (loading, unloading, rest, etc.) without integrating small parasite stops (congestion, traffic lights, small OBU distortion, etc.). Using another time threshold would obviously lead to results that may be drastically different and corresponding to others freight realities. Incorporating land uses information would help to control the nature of the stops.

Finally, many cleaning steps and filters were applied to the dataset. Cleaning the data obviously leads to a better control of this large amount of information and avoids considering irregular effects. Nevertheless, this methodological framework erases more or less 30% of the initial information, questioning us about the final organisation of the dataset. Do we finally measure the initial movements or do the cleaning steps lead to transform the nature of the data on what we are looking for?

8.3 Paths for further research

From the outputs, several unsolved questions were raised and this Section proposes to trace paths for further research. Section 8.3.1 focusses on some methodological aspects, while Section 8.3.2 explores more deeply mobile phone calls. Finally, section 8.3.3 takes the road for future analyses of the freight exchanges based on the spatio-temporal following of trucks.

8.3.1 Linked with community detection method

Community detection methods are useful to detect groups of nodes that are highly connected within a network. Applied to a geographical context, and without any knowledge of the location of the nodes, these methods group contiguous places together, proving the strong effect of distance on movements of people, freight and information. Close places interact more than distant places (Tobler, 1970).

An interesting question would be to fully integrate geographical insights in community detection methods by giving characteristics to the nodes. This principle was already initiated by Newman (2003) with the development of the assortativity, that is a preference for a network's node to attach to others that are similar in some way. By using well-known characteristics of nodes (socio-economic information, spatial hierarchy, etc.), classification into communities would be done by maximising two different objectives: classifying together places that 1) are similar in terms of characteristics and 2) have a strong interrelation.

Another challenging research would be to consider the effect of distance in community detection methods. By giving more weight to interaction made on longer distance, do we drastically modify the spatial structure? Could results be linked with the spatial interaction models (Section 1.2.2) and more particularly the well-known gravity model improved by Wilson (1974)? This kind of question was initiated by Lhomme (2017), but the interpretation of the results are difficult, while considering the weight given attributed to the distance. Another approach was choose by Beckers *et al.* (2019). Instead of only classify nodes with community detection, these authors decide to consider borders as major geographical features by quantifying them though damping values. These damping values are mainly based on a hierarchical distance, expressing that node classified in a same community, and at a same level of resolution are close to each other. With this combination, Beckers *et al.* (2019) succeed to derive a certain hierarchical structure in the case of logistics buyer-supplier network in Belgium. However, this contribution is approximating distance (called hierarchical distance) based on the assumption that communities are always composed of contiguous places, which is not always true has shown by the communities of antennas presented in this thesis (Chapter 7). These show that although steps are made in this direction, improvements in the theoretical and conceptual framework has to be made to better combine distance and interaction in community detection method.

This thesis partially combines geographical features and community detection methods in the context of transport geography: interaction fields detected by community detection methods were compared to urban models used as a proxy of the characteristics. However, the distance between places was not considered in our analyses. Future works are required to fully integrate interaction, characteristics and distance, proposing new ways to model transport geography in the era of big-data.

8.3.2 Linked with mobile phone calls

Analyses made on mobile phone calls highlight movements of information through space, as well as social interactions between people and between their places of life. Unfortunately, our analyses were made on a sample of users who only communicate within the Former province of Brabant. A smart challenge would be to put together calls, text messages and CRD details, whatever the mobile phone company. This would obviously lead to an accurate and exhaustive view on social interactions of people owning mobile phones. However, this would raise several issues linked to the incompatibility of these various datasets and with respect condition of life privacy.

8.3.3 Linked with the real-time following of trucks

Analyses applied to the movements of trucks are anchorage in a current (endless) debate: should we refer to *data-driven* or *theory-driven* models? We decided to refer to *data-driven* approach, and more particularly on the *Geographic knowledge-discovery*. This theoretical approach refers to "*the initial stage of the scientific process where the investigator forms his or her conceptual view of the system, develops hypotheses to be tested, and performs groundwork to support the knowledge-construction process*" (Miller and Goodchild, 2015). Based on a description of the database and its structure, we conceptualise a framework of cleaning and adaptation to reach our objectives (Section 6.3.1). Nevertheless, analyses were only applied to our objectives without any considerations of the roads used or the travel time. Further research should be concentrated on the spatial organisation of the movements made by truck, in order to enrich our *data-driven* model by hypothesis and observations, leading to a more *theory-driven* model. Amongst other, further research should investigate the properties of the stops (land use characteristics, economic activities) as well as those of the edges (length, type of roads, traffic volume). What additional information can we get about the spatio-temporal organisation of supply chains when looking at truck routes (as a succession of nodes and edges)? Is it possible to classify route shapes?

8.4 Implications and recommendations

About commuting and residential moves, the necessity of a Metropolitan management for mobility and housings questions

As already explained in Section 8.1.3, commuting and migration movements have very specific spatial footprints that do not accurately followed the

delineation of the BCR. The demographic and economic development of Brussels should not be hindered by administrative levels that not fit with the "real" urban extension, as seen by the workers or by the residents. Policies of land planning should more largely consider this inconsistency between the administrative level and the real urban extension of Brussels, sprawling on the Flemish and Walloon Brabant. Although this topic is sensitive, it could be more accurate that studies, or land planning, directly integrate the Metropolitan scale instead of only the regional one. A strong metropolitan institution would hence be highly welcomed to manage policies linked with housings or mobility. Territory development or urban planning should consider the urban envelope as a whole.

About train schedules requests, the possibility of the Brussels region to identify and answer the users' needs

Combining the train schedules requests with the real user number of major public transportation companies (SNCB and STIB), would highlight the transportation demands between places. Temporal dimension should also integrated with a high level of accuracy. This data collection will help to evaluate what should be the location of a new specific infrastructure offering to the population the highest propensity to travel by train (or other kind of public transport). It will also help to improve timetables that maximise the use of public transportation services. Hence, the Region will have at disposal an original and exhaustive platform that will accurately answer to the demand for transportation made by travellers.

About phone communication, the possibility to improve census and urban planning

With the new era raised by big-data, official statistical services should be aware of potentialities coming out of crossing census and mobile phone data. Statistics Belgium (Debusschere *et al.*, 2017), as well as other official statistics services of foreign countries (Sakarovitch *et al.*, 2018; Scholtus, 2015), works on these approaches and already developed useful frameworks. We can easily imagine that using mobile phone datasets as complementary information would improve the data provided by the census by taking the advantage of both datasets. To collect an exhaustive information of positions of each people, as well as their uses on the phone network (calls, text messages, internet connection), data from every mobile phone operators should be collected.

All along this project, the classification into communities of the tightly connected antennas was made on several levels of resolution (Section 2.2.1),

leading to detect a varying number of groups. This multi-level approach leads to a better understanding of how the various places are organised between them. Sometime, the Louvain Method leads to communities always detected through different values of ρ. These particular places, hence highly connected between them, have a small propensity to interact with the rest of the Metropolis. These isolated groups of places may be explained by the spatial segregation observed in the BCR (Section 1.1 and Van Hamme *et al.*, 2016). Only few information exists (based on surveys or samples) concerning the real communication patterns (and hence the social networks) between the numerous places of life within Brussels. A better understanding of these communications is useful for land planning purposes in order to correctly integrate these segregated places within a strong Brussels "community".

About the circulation of trucks, the possibility to elaborate sustainable policies

For many years, the road mobility around Brussels is a major negative aspect. Diverse studies based on local counting points or surveys (not exhaustive by definition) were conducted to punctually estimate the flow of vehicles or the drivers' choices. The Viapass dataset passes through the issues linked with samples by proposing information for each vehicle of more than 3,5 tons driving in Belgium. Using a simple automatic implementation, it is possible to create a real-time monitoring software of the circulation of trucks that highlights congestion areas and that could be used for decision-making. In addition, this tool would be helpful to deviate the upstream traffic to other axes when congestion/accident are observed, leading to a regulation of the general traffic. The automatic sending of emergency services is also a major opportunity. A main advantage of the Viapass dataset, as regards to other GPS data sources (TomTom, Coyote, Google Maps), resides in the following of trucks that are driving earlier than other vehicles. By analysing the circulation of trucks, it is easy to detect any issues appearing on the road network before the major flow of vehicles starts (and hence before the detection by TomTom, Coyote or Google).

As a parallel aspect of the traffic, understanding areas attracting trucks is an important knowledge. Indeed, more than just regulate the traffic, evaluate the stopping areas as well as their average duration will be the basis for the creation of catchment areas reachable during a specific time period (or forbidden during certain peak periods), like it is already developed in the centre of Brussels. Communicate these attendance information to road hauliers will help them to optimise their daily delivery circuit by considering the traffic saturation around these major places. Because small vans of less than 3.5 tons are important in urban areas, these policies should include these vehicles. By extending works

and adding complementary analyses, it should be interesting to discuss the pertinence and the implementation of a public system that manages the "last kilometre" between the Ring of Brussels and the destinations of trucks. With the knowledge of the features of each truck, it becomes simpler to evaluate and anticipate the number of lorries considered, as well as the major axes used. Within the current framework restricting accessibility to the BCR for the most polluting vehicles, this kind of system will hence help to develop a strong legislative accessibility framework of every vehicle. More generally, this would drastically reduce the negative externalities induced by vehicles of more than 3.5 tons: amongst other pollution, congestion (linked by truck stop double-parked for unloading) or the impact on roads.

About the use of big-datasets, the importance of good practices

Whatever a dataset come from censuses or sensors, a particular attention should be done on the understanding, cleaning and storage. There is always a need for processing steps that transform raw data into a smart usable information. These processes were applied by statistics services, leading to a great understanding of the biases that may be included. However, researchers dealing with big-data have to do these cleaning steps by themselves, but without a strong global theoretical framework, the information created is specific of their own research (Miller and Goodchild, 2015), as show in this thesis. Cleaning steps developed for analysing the Viapass dataset were costly (Section 6.2.2), but the computed information answer our objectives. We hope that the framework designed will be used by other researcher dealing with movements of trucks.

The representativeness is an important aspect of data. As shown in this thesis, mobile phone calls lead to partition space into well-delineated areas. However, we have no information related to calls nor the characteristics of the people: we hence analysed calls made between unknown people for unknown reasons based on an unknown sample of the population. It shows that although we made considerable technological advances for analysing interaction patterns, we still have too much uncertainty. Can we really extrapolate our results to social sciences, urban planers or policy makers? Beside this specific case, representativeness and exhaustiveness could be found in ICT datasets, like in the Viapass dataset that mostly contains the moves of mostly all trucks circulating in Belgium.

These show that big-data are like any databases. Although, they have to be controlled to avoid any inherent biases, they are useful to bring local information in our more global model (Batty, 2012). However, in the current era of *data-driven* models, there is a crucial need for a strong theoretical

framework making our cleaning steps more transparent and reusable in similar context.

8.5 Concluding words

The objective of this thesis was to measure, understand and explain the interrelations observed in and around the metropolitan area of Brussels by analysing networks of movements.

Instead of using spatial interaction methods, analyses of interactions were made by the Louvain Method. This community approach, not concerned about the characteristics of the places, shows how interaction observed between people/places are organised, and embedded into several levels of resolution. This framework reveals that studying space through interactions is not trivial. Using several datasets different per nature leads to results that are difficultly combinable for offering a unique outcome.

However, by considering spatial big-datasets, this thesis provides insights about local interaction appearing during specific period of the time. ICT data hence adds short-term information in our more long-term and global models. It ends up with new views on the spatial organisation of the Belgian territory, as well as of the BCR. A better combination of these datasets with censuses would considerably increase the knowledge we already have about our places of life.

By combining characteristics of the places with the interaction fields, this thesis shows that transport geography theory is a major theoretical framework that should be considered in the current era of data-driven models. However, it hence imposes a better integration of geographical notions, like distance and characteristics, into community detection methods, leading to new ways of studying transport of people, freight and information in space.

Bibliography

Adam, A., Charlier, J., Debuisson, M., Duprez, J.P., Reginster, I., and Thomas, I. (2018a). "Residential basins in Belgium: Two methods, one reality?" *L'Espace géographique*, Volume 47(1): 35–50. URL https://www.cairn-int.info/article.php?ID_ARTICLE=E_EG_471_0035&USER=a.adam@uclouvain.be.

Adam, A., Delvenne, J.C., and Thomas, I. (2017). "Cartography of interaction fields in and around Brussels: commuting, moves and telephone calls". *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles / Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels.* URL http://journals.openedition.org/brussels/1601.

Adam, A., Delvenne, J.C., and Thomas, I. (2018b). "Detecting communities with the multi-scale Louvain method: robustness test on the metropolitan area of Brussels". *Journal of Geographical Systems*, 20(4): 363–386. URL https://doi.org/10.1007/s10109-018-0279-0.

Aggarwal, C.C. (2014). *Data classification: algorithms and applications.* CRC press.

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., and Zook, M. (2015). "Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn". *International Journal of Geographical Information Science*, 29(11): 2017–2039. URL http://www.tandfonline.com/doi/abs/10.1080/13658816.2015.1063151.

Ana, L.N.F. and Jain, A.K. (2003). "Robust data clustering". In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pp. II–II.

Anas, A., Arnott, R., and Small, K. (1998). "Urban Spatial Structure". *Journal of Economic Literature*, 36(3): 1426–1464. URL http://econpapers.repec.org/article/aeajeclit/v_3a36_3ay_3a1998_3ai_3a3_3ap_3a1426-1464.htm.

Andresen, M.A. (2010). "Geographies of International Trade: Theory, Borders, and Regions". *Geography Compass*, 4(2): 94–105. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2009.00314.x.

Antoniou, C., Balakrishna, R., and Koutsopoulos, H.N. (2011). "A Synthesis of emerging data collection technologies and their impact on traffic management applications". *European Transport Research Review*, 3(3): 139–148. URL https://doi.org/10.1007/s12544-011-0058-1.

Arenas, A., Fernández, A., and Gómez, S. (2008). "Analysis of the structure of complex networks at different resolution levels". *New Journal of Physics*, 10(5): 053 039. URL http://stacks.iop.org/1367-2630/10/i=5/a=053039.

Attout, X. (2016). "RER: scenario catastrophe en vue?" *Le Soir*.

Bacaër, N. (2011). "Verhulst and the logistic equation (1838)". In *A Short History of Mathematical Population Dynamics*, pp. 35–39. Springer, London. URL https://link.springer.com/chapter/10.1007/978-0-85729-115-8_6.

Bailey, T. and Gatrell, A. (1995). "Interactive spatial data analysis". URL http://www.personal.psu.edu/faculty/f/k/fkw/rsoc597/Introduction.pdf.

Barthelemy, M. (2016). *The structure and dynamics of cities*. Cambridge University Press.

Batty, M. (2012). "Smart Cities, Big Data". *Environment and Planning B: Planning and Design*, 39(2): 191–193. URL http://epb.sagepub.com/content/39/2/191.

Batty, M. (2013). *The New Science of Cities*. MIT Press.

Batty, M. and Torrens, P.M. (2001). "Modelling complexity : The limits to prediction". *Cybergeo : European Journal of Geography*. URL http://journals.openedition.org/cybergeo/1035.

Beckers, J., Thomas, I., Vanoutrive, T., and Verhetsel, A. (2018). "Logistics clusters, including inter-firm relations through community detection". *European Journal of Transport and Infrastructure Research*,

18(2). URL https://journals.open.tudelft.nl/index.php/ejtir/article/view/3229.

Beckers, J., Vanhoof, M., and Verhetsel, A. (2019). "Returning the particular: Understanding hierarchies in the Belgian logistics system". *Journal of Transport Geography*, 76: 315–324. URL http://www.sciencedirect.com/science/article/pii/S0966692316307645.

Beguerisse-Díaz, M., Garduño-Hernández, G., Vangelov, B., Yaliraki, S.N., and Barahona, M. (2014). "Interest communities and flow roles in directed networks: the Twitter network of the UK riots". *Journal of The Royal Society Interface*, 11(101): 20140 940. URL http://classic.rsif.royalsocietypublishing.org/content/11/101/20140940.short.

Bergstrand, J.H. (1985). "The gravity equation in international trade: some microeconomic foundations and empirical evidence". *The review of economics and statistics*, pp. 474–481.

Blondel, V., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10 008. URL http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52.

Blondel, V., Krings, G., and Thomas, I. (2010). "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone". *Brussels Studies*, 42(4): 1–12. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.187.3791&rep=rep1&type=pdf.

Boussauw, K., Allaert, G., and Witlox, F. (2013). "Colouring Inside What Lines? Interference of the Urban Growth Boundary and the Political–Administrative Border of Brussels". *European Planning Studies*, 21(10): 1509–1527. URL https://doi.org/10.1080/09654313.2012.722952.

Boyd, d. and Crawford, K. (2012). "Critical Questions for Big Data". *Information, Communication & Society*, 15(5): 662–679. URL http://dx.doi.org/10.1080/1369118X.2012.678878.

Burgess, E.W. (1925). *The growth of the city: an introduction to a research project. The City.* University of Chicago Press, Chicago.

Calabrese, F., Lorenzo, G.D., Liu, L., and Ratti, C. (2011). "Estimating Origin-Destination Flows Using Mobile Phone Location Data". *IEEE Pervasive Computing*, 10(4): 36–44.

Carter, H. (1981). *The study of urban geography.*

Census-2011 (2016). "Census 2011". URL http://census2011.fgov.be/index_fr.html.

Cerina, F., Chessa, A., Pammolli, F., and Riccaboni, M. (2014). "Network communities within and across borders". *Scientific Reports*, 4. URL http://www.nature.com/doifinder/10.1038/srep04546.

Cervero, R. and Wu, K.L. (1997). "Polycentrism, commuting, and residential location in the San Francisco Bay area". *Environment and Planning A*, 29(5): 865–886.

Charlier, J., Debuisson, M., Duprez, J.P., and Reginster, I. (2016). "Mouvements résidentiels en Wallonie (1994-2014): analyses des migrations intercommunales et construction de bassins résidentiels". Working paper 21, IWEPS, Namur. URL http://www.iweps.be/sites/default/files/wp21.pdf.

Cheshire, P. (2010). "Why Brussels needs a City-Region for the City". *Re-bel ebook 7 What does Geography teach us about the future of the Belgium's Institution ?*, pp. 33–48. URL http://www.rethinkingbelgium.eu/rebel-initiative-files/ebooks/ebook-7/Cheshire.pdf.

Chi, G., Thill, J.C., Tong, D., Shi, L., and Liu, Y. (2016). "Uncovering regional characteristics from mobile phone data: A network science approach". *Papers in Regional Science*, 95(3): 613–631. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/pirs.12149.

Clerval, A. and Van Criekingen, M. (2014). "«Gentrification ou ghetto», décryptage d'une impasse intellectuelle". *Métropolitiques. eu.* URL https://www.metropolitiques.eu/IMG/pdf/pdf_met-clerval_et_van_criekingen.pdf.

Commenges, H. (2013). *L'invention de la mobilité quotidienne. Aspects performatifs des instruments de la socio-économie des transports.* phdthesis, Université Paris-Diderot - Paris VII. URL https://tel.archives-ouvertes.fr/tel-00923682/document.

Coombes, M. (2013). "From City-region Concept to Boundaries for Governance: The English Case". *Urban Studies.*

Cottineau, C. and Vanhoof, M. (2019). "Mobile Phone Indicators and Their Relation to the Socioeconomic Organisation of Cities". *ISPRS International Journal of Geo-Information*, 8(1): 19. URL https://www.mdpi.com/2220-9964/8/1/19.

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., and Zook, M. (2013). "Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb". *Cartography and Geographic Information Science*, 40(2): 130–139. URL http://www.tandfonline.com/doi/abs/10.1080/15230406.2013.777137.

Cuzzocrea, A., Song, I.Y., and Davis, K.C. (2011). "Analytics over large-scale multidimensional data: the big data revolution!" In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP*, pp. 101–104. ACM. URL http://dl.acm.org/citation.cfm?id=2064695.

Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005). "Comparing community structure identification". *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09): P09 008. URL http://iopscience.iop.org/1742-5468/2005/09/P09008.

De Keersmaecker, M.L., Frankhauser, P., and Thomas, I. (2003). "Using Fractal Dimensions for Characterizing Intra-urban Diversity: The Example of Brussels". *Geographical Analysis*, 35(4): 310–328. URL http://dx.doi.org/10.1111/j.1538-4632.2003.tb01117.x.

De Maesschalck, F., De Rijck, T., and Heylen, V. (2015). "Au-delà de la frontière : relations socio-spatiales entre Bruxelles et le Brabant flamand". *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles / Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels*. URL http://brussels.revues.org/1259.

de Wasseige, Y., Laffut, M., and Ruyters, C. (2000). *Bassins d'emploi et régions fonctionnelles: méthodologie et définition des bassins d'emploi belges*. Ministère de la région wallonne. Service des études et de la statistique.

Deboosere, P., Eggerickx, T., Van Hecke, E., Wayens, B., and others (2009). "La population bruxelloise: un éclairage démographique". *Brussels Studies*, pp. 1–18. URL http://www.brusselsstudies.be/medias/publications/FR_71_EGB3.pdf.

Debusschere, M., Lusyne, P., Dewitte, P., Baeyens, Y., Meersman, F.D., Seynaeve, G., Demunter, C., Reis, F., and Reuter, H.I. (2017). "Big data et statistiques : un recensement tous les quarts d'heure...". URL https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/analyse_bigdata_fr.pdf.

Decuyper, A., Gandica, Y., Cloquet, C., Thomas, I., and Delvenne, J.C. (2018). "Measuring the effect of node aggregation on community

detection". *arXiv:1809.08855 [physics]*. URL http://arxiv.org/abs/1809.08855, arXiv: 1809.08855.

Delloye, J. (2018). *Urban Morphodynamics*. phdthesis, Université catholique de Louvain (Louvain-la-Neuve, Belgique). URL https://tel.archives-ouvertes.fr/tel-01847861/document.

Delmotte, A., Tate, E.W., Yaliraki, S.N., and Barahona, M. (2011). "Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin–myosin light chain interaction". *Physical Biology*, 8(5): 055 010. URL http://stacks.iop.org/1478-3975/8/i=5/a=055010.

Delvaux, B. (2005). "Méthode de définition des espaces d'interdépendance entre écoles". Technical Report 24, Louvain la Neuve.

Delvenne, J.C., Schaub, M.T., Yaliraki, S.N., and Barahona, M. (2013). "The Stability of a Graph Partition: A Dynamics-Based Framework for Community Detection". In *Dynamics On and Of Complex Networks, Volume 2*, Modeling and Simulation in Science, Engineering and Technology, pp. 221–242. Birkhäuser, New York, NY. URL https://link.springer.com/chapter/10.1007/978-1-4614-6729-8_11.

Delvenne, J.C., Yaliraki, S.N., and Barahona, M. (2010). "Stability of graph communities across time scales". *Proceedings of the National Academy of Sciences*, 107(29): 12 755–12 760. URL http://www.pnas.org/content/107/29/12755.

Denis, J (1992). *Geographie de la Belgique*. Bulletin du Credit Communal, Brussels, Belgium.

Der Herten, V. (2001). *Le temps du train: 175 ans de chemins de fer en Belgique, 75e anniversaire de la SNCB*. Presses Universitaires de Louvain. URL https://books.google.be/books?id=6j1KAAAACAAJ.

Dessouroux, C., Van Criekingen, M., and Decroly, J.M. (2009). "Embellissement sous surveillance: une géographie des politiques de réaménagement des espaces publics au centre de Bruxelles". *Belgeo. Revue belge de géographie*, (2): 169–186. URL http://belgeo.revues.org/7946.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., and Tatem, A.J. (2014). "Dynamic population mapping using mobile phone data". *Proceedings of the National Academy of Sciences*, 111(45): 15 888–15 893. URL http://www.pnas.org/content/111/45/15888.

Dobruszkes, F. (2012). "Stimulating or frustrating research ? Transport geography and (un)available data". *Belgeo. Revue belge de géographie*, (1-2). URL http://journals.openedition.org/belgeo/7082.

Dobruszkes, F. and Vandermotten, C. (2006). "Éléments pour une géographie des clivages philosophiques à Bruxelles". *LEspace geographique*, Tome 35(1): 31–43. URL https://www.cairn.info/revue-espace-geographique-2006-1-page-31.htm.

Dujardin, C. (2001). "Effet de frontière et interaction spatiale". *L'Espace géographique*, tome 30(4): 307–320. URL https://www.cairn.info/revue-espace-geographique-2001-4-page-307.htm.

Dujardin, C., Selod, H., and Thomas, I. (2004). "Le chômage dans l'agglomération bruxelloise: une explication par la structure urbaine". *Revue d'Économie Régionale & Urbaine*, février(1): 3–27.

Dujardin, C., Selod, H., and Thomas, I. (2008). "Residential Segregation and Unemployment: The Case of Brussels". *Urban Studies*, 45(1): 89–113. URL http://usj.sagepub.com/cgi/doi/10.1177/0042098007085103.

Dujardin, C., Thomas, I., and Tulkens, H. (2007). "Quelles frontières pour Bruxelles? Une mise à jour". *Reflets et perspectives de la vie économique*, XLVI(2-3): 155–176.

Duranton, G. (2015). "Delineating Metropolitan Areas: Measuring Spatial Labour Market Networks Through Commuting Patterns". In T. Watanabe, I. Uesugi, and A. Ono, editors, *The Economics of Interfirm Networks*, number 4 in Advances in Japanese Business and Economics, pp. 107–133. Springer Japan. URL http://link.springer.com/chapter/10.1007/978-4-431-55390-8_6.

Eggerickx, T. (2013). "Transition démographique et banlieue en Belgique: le cas de Bruxelles". *Annales de démographie historique*, 126(2): 51–80.

Eggerickx, T., Hermia, J.P., Surkijn, J., and Willaert, D. (2010). "Les migrations internes en Belgique". URL https://dial.uclouvain.be/pr/boreal/object/boreal:78957.

Ermans, T., Brandeleer, C., Hubert, M., Lebrun, K., and Sieux, F. (2018). "Déplacements domicile-travail : état des lieux et perspectives d'action pour les entreprises. Note de synthèse BSI". *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles / Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels.* URL http://journals.openedition.org/brussels/1680.

Estivill-Castro, V. (2002). "Why So Many Clustering Algorithms: A Position Paper". *SIGKDD Explor. Newsl.*, 4(1): 65–75. URL http://doi.acm.org/10.1145/568574.568575.

Eurostat (2016). "Geostat". URL https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/population-distribution-demography/geostat.

Expert, P., Evans, T.S., Blondel, V.D., and Lambiotte, R. (2011). "Uncovering space-independent communities in spatial networks". *Proceedings of the National Academy of Sciences*, 108(19): 7663–7668. URL http://www.pnas.org/content/108/19/7663.short.

Fan, W. and Bifet, A. (2013). "Mining big data: current status, and forecast to the future". *ACM SIGKDD Explorations Newsletter*, 14(2): 1–5. URL http://dl.acm.org/citation.cfm?id=2481246.

Finance, O., Adam, A., Jones, J., and Thomas, I. (2018). "Révéler la polarisation économique d'une ville à partir de traces GPS de camions. Le cas de Liège". URL https://dial.uclouvain.be/pr/boreal/object/boreal:209776.

Fischer, M.M., Essletzbichler, J., Gassler, H., and Trichtl, G. (1993). "Telephone Communication Patterns in Austria: A Comparison of the IPFP-based Graph-Theoretic and the Intramax Approaches". *Geographical Analysis*, 25(3): 224–233. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1993.tb00293.x/abstract.

Flaskou, M., Dulebenets, M.A., Golias, M.M., Mishra, S., and Rock, R.M. (2015). "Analysis of Freight Corridors Using GPS Data on Trucks". *Transportation Research Record: Journal of the Transportation Research Board*, 2478: 113–122. URL https://trrjournalonline.trb.org/doi/abs/10.3141/2478-13.

Fortunato, S. (2009). "Community detection in graphs". *Physics Reports*, 486(3-5): 75–174. URL http://linkinghub.elsevier.com/retrieve/pii/S0370157309002841.

Fortunato, S. and Barthélemy, M. (2007). "Resolution limit in community detection". *Proceedings of the National Academy of Sciences*, 104(1): 36–41. URL http://www.pnas.org/content/104/1/36.

Fortunato, S. and Lancichinetti, A. (2009). "Community detection algorithms: a comparative analysis: invited presentation, extended abstract". In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, p. 27. ICST (Institute for Computer

Sciences, Social-Informatics and Telecommunications Engineering). URL http://dl.acm.org/citation.cfm?id=1698858.

Fotheringham, A.S. (2001). "Spatial Interaction Models". In N.J. Smelser and P.B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pp. 14 794–14 800. Pergamon, Oxford. URL http://www.sciencedirect.com/science/article/pii/B0080430767025195.

Fotheringham, A.S. and O'Kelly, M.E. (1989). *Spatial interaction models: formulations and applications*. Kluwer Academic Dordrecht.

Freeman, L.C. (1977). "A set of measures of centrality based on betweenness". *Sociometry*, pp. 35–41.

Freeman, L.C. (1978). "Centrality in social networks conceptual clarification". *Social Networks*, 1(3): 215–239. URL http://www.sciencedirect.com/science/article/pii/0378873378900217.

Fujita, M. and Thisse, J.F. (2013). *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge University Press.

Furletti, B., Trasarti, R., Cintia, P., and Gabrielli, L. (2017). "Discovering and Understanding City Events with Big Data: The Case of Rome". *Information*, 8(3): 74. URL https://www.mdpi.com/2078-2489/8/3/74.

Gabrielli, L., Furletti, B., Trasarti, R., Giannotti, F., and Pedreschi, D. (2015). "City users' classification with mobile phone data". In *2015 IEEE International Conference on Big Data (Big Data)*, pp. 1007–1012.

Gariazzo, C., Pelliccioni, A., and Bogliolo, M.P. (2019). "Spatiotemporal Analysis of Urban Mobility Using Aggregate Mobile Phone Derived Presence and Demographic Data: A Case Study in the City of Rome, Italy". *Data*, 4(1): 8. URL https://www.mdpi.com/2306-5729/4/1/8.

Gaspar, J. and Glaeser, E.L. (1998). "Information Technology and the Future of Cities". *Journal of Urban Economics*, 43(1): 136–156. URL http://www.sciencedirect.com/science/article/pii/S0094119096920318.

Gingerich, K., Maoh, H., and Anderson, W. (2016). "Classifying the purpose of stopped truck events: An application of entropy to GPS data". *Transportation Research Part C: Emerging Technologies*, 64: 17–27. URL http://www.sciencedirect.com/science/article/pii/S0968090X16000036.

Girvan, M. and Newman, M.E.J. (2002). "Community structure in social and biological networks". *Proceedings of the National Academy of Sciences*, 99(12): 7821–7826. URL http://www.pnas.org/content/99/12/7821.

Givoni, M. (2006). "Development and impact of the modern High-speed train: A review". *Transport reviews*, 26(5): 593–611.

Glaeser, E. (2011). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier and Happier*. Pan Macmillan.

Géoconfluences (2017). "Vocabulaire et notions générales". URL http://geoconfluences.ens-lyon.fr/glossaire.

González, M.C., Hidalgo, C.A., and Barabási, A.L. (2008). "Understanding individual human mobility patterns". *Nature*, 453(7196): 779–782. URL http://www.nature.com/doifinder/10.1038/nature06958.

Goodchild, M.F. (2013). "The quality of big (geo)data". *Dialogues in Human Geography*, 3(3): 280–284. URL https://doi.org/10.1177/2043820613513392.

Gottschalk, P.G. and Dunn, J.R. (2005). "The five-parameter logistic: A characterization and comparison with the four-parameter logistic". *Analytical Biochemistry*, 343(1): 54–65. URL http://www.sciencedirect.com/science/article/pii/S0003269705003313.

Gower, J.C. (1971). "A General Coefficient of Similarity and Some of Its Properties". *Biometrics*, 27(4): 857–871. URL http://www.jstor.org/stable/2528823.

Graells-Garrido, E., Peredo, O., and García, J. (2016). "Sensing urban patterns with antenna mappings: the case of Santiago, Chile". *Sensors*, 16(7): 1098.

Graham, M. and Shelton, T. (2013). "Geography and the future of big data, big data and the future of geography". *Dialogues in Human Geography*, 3(3): 255–261. URL http://dhg.sagepub.com/lookup/doi/10.1177/2043820613513121.

Grasland, C. (1999). "Interaction spatiale et effets de barrière. 1re partie : Des modèles d'interaction spatiale aux modèles d'interaction territoriale". *CNRS-UMR Géographie-Cités*.

Grasland, C. and Beauguitte, L. (2010). "Modelling attractiveness of global places: A worldwide survey on 9000 undergraduate students". In *50th European Congress of the Regional Science Association International*

on: Sustainable Regional Growth and Development in the Creative Knowledge Economy-19th to 23th of August.

Guimera, R., Mossa, S., Turtschi, A., and Amaral, L.N. (2005). "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles". *Proceedings of the National Academy of Sciences*, 102(22): 7794–7799.

Hall, T., Barrett, H., and Barrett, H. (2012). *Urban Geography*. Routledge. URL https://www.taylorfrancis.com/books/9781136647369.

Hao, J., Zhu, J., and Zhong, R. (2015). "The rise of big data on urban studies and planning practices in China: Review and open research issues". *Journal of Urban Management*, 4(2): 92–124.

Harris, C.D. and Ullman, E.L. (1945). "The Nature of Cities". *The Annals of the American Academy of Political and Social Science*, 242: 7–17. URL http://www.jstor.org/stable/1026055.

Haynes, K.E. and Fotheringham, A.S. (1984). *Gravity and spatial interaction models*, volume 2. Sage publications Beverly Hills. URL http://www.web.pdx.edu/~stipakb/download/PA557/ReadingsPA557sec1-2.pdf.

Herman, R.A., Scherer, P.N., and Shan, G. (2008). "Evaluation of logistic and polynomial models for fitting sandwich-ELISA calibration curves". *Journal of Immunological Methods*, 339(2): 245–258. URL http://www.sciencedirect.com/science/article/pii/S0022175908002809.

Hermia, J.P. (2005). "Une nouvelle délimitation spatiale du phénomène périurbain bruxellois". *Union Internationale pour l'Etude Scientifique de la Population. XXVe Congrès International de la Population., Tours, France*, pp. 18–23. URL http://demoscope.ru/weekly/knigi/tours_2005/papers/iussp2005s51904.pdf.

Horner, M.W. and Murray, A.T. (2002). "Excess Commuting and the Modifiable Areal Unit Problem". *Urban Studies*, 39(1): 131–139. URL http://usj.sagepub.com/content/39/1/131.

Hoyt, H. (1939). "The structure and growth of residential neighborhoods in American cities". *journal*.

Huff, D.L. (1963). "A probabilistic analysis of shopping center trade areas". *Land economics*, 39(1): 81–90.

Infrabel (2019). "Railway lines". URL http://www.infrabel.be/en/about/our-rail-network/railway-lines.

insee (2015). "Les zonages d'étude de l'Insee: une histoire des zonages supracommunaux définis à des fins statistiques". *Insee méthode*, p. 97.

Ioannides, Y.M. (2012). "From Neighborhoods to Nations: The Economics of Social Interactions". Economics Books, Princeton University Press. URL https://ideas.repec.org/b/pup/pbooks/9892.html.

Jacobs, J. (1969). *The economy of cities*. Random House, New-York.

Janson, C.G. (1980). "Factorial social ecology: An attempt at summary and evaluation". *Annual Review of Sociology*, 6(1): 433–456.

Jones, J., Cloquet, C., Adam, A., Decuyper, A., and Thomas, I. (2016). "Belgium through the Lens of Rail Travel Requests: Does Geography Still Matter?" *ISPRS International Journal of Geo-Information*, 5(11): 216. URL http://www.mdpi.com/2220-9964/5/11/216.

Jones, J., Peeters, D., and Thomas, I. (2017). "Scale effect in a LUTI model of Brussels: challenges for policy evaluation". *EJTIR*, 17(1): 103–131. URL http://www.tbm.tudelft.nl/fileadmin/Faculteit/TBM/Onderzoek/EJTIR/Back_issues/17.1/2017_01_05.pdf.

Joubert, J.W. and Meintjes, S. (2015). "Repeatability & reproducibility: Implications of using GPS data for freight activity chains". *Transportation Research Part B: Methodological*, 76: 81–92. URL http://www.sciencedirect.com/science/article/pii/S0191261515000508.

Järv, O., Müürisepp, K., Ahas, R., Derudder, B., and Witlox, F. (2014). "Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia". *Urban Studies*, p. 0042098014550459. URL http://usj.sagepub.com/content/early/2014/09/22/0042098014550459.

Järv, O., Tenkanen, H., Salonen, M., Ahas, R., and Toivonen, T. (2018). "Dynamic cities: Location-based accessibility modelling as a function of time". *Applied Geography*, 95: 101–110. URL https://www.sciencedirect.com/science/article/pii/S014362281731144X.

Kesteloot, C., Meert, H., and Mistiaen, P. (1995). "Polarisation socio-spatiale et stratégies de survie dans deux quartiers bruxellois." *Espace, populations, sociétés*, pp. 277–290. URL /web/revues/home/prescript/article/espos_0755-7809_1995_num_13_3_1703.

Ketchen, D.J. and Shook, C.L. (1996). "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique". *Strategic Management Journal*, 17(6): 441–458. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0266%28199606%2917%3A6%3C441%3A%3AAID-SMJ819%3E3.0.CO%3B2-G.

Kitchin, R. (2013). "Big data and human geography: Opportunities, challenges and risks". *Dialogues in Human Geography*, 3(3): 262–267. URL http://dhg.sagepub.com/lookup/doi/10.1177/2043820613513388.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE.

Kitchin, R. and Tate, N. (1999). *Conducting Research in Human Geography: theory, methodology and practice*. Longman, 1 edition.

Kraak, M. and Ormeling, F. (2011). *Cartography visualization of spatial data*. Guildford Press., New York.

Kupfer, F., Kessels, R., Goos, P., Van de Voorde, E., and Verhetsel, A. (2016). "The origin–destination airport choice for all-cargo aircraft operations in Europe". *Transportation Research Part E: Logistics and Transportation Review*, 87: 53–74. URL http://www.sciencedirect.com/science/article/pii/S1366554515002240.

Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L., and Nippani, S. (2014). "Development of a tour-based truck travel demand model using truck GPS data". Technical report.

Lambiotte, R. (2010). "Multi-scale modularity in complex networks". In *Modeling and optimization in mobile, ad hoc and wireless networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*, pp. 546–553. IEEE. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5520347.

Lambiotte, R., Delvenne, J.C., and Barahona, M. (2014). "Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks". *IEEE Transactions on Network Science and Engineering*, 1(2): 76–90.

Laurila, J.K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J., Miettinen, M., and others (2012). "The mobile data challenge: Big data for mobile computing research". In *Pervasive Computing*. URL http://infoscience.epfl.ch/record/192489.

Lawlor, A., Coffey, C., McGrath, R., and Pozdnoukhov, A. (2012). "Stratification structure of urban habitats". p. 8.

Lebowitz, M.D. (1977). "A Critical Examination of Factorial Ecology and Social Area Analysis for Epidemiological Research". *Journal of the Arizona Academy of Science*, 12(2): 86–90. URL https://www.jstor.org/stable/40022220.

Lee, J.G. and Kang, M. (2015). "Geospatial Big Data: Challenges and Opportunities". *Big Data Research*. URL http://www.sciencedirect.com/science/article/pii/S2214579615000040.

Leemans, S., Pattyn, M., Rousseau, S., and Van der Haegen, H. (1989). "De Belgische stadsgewesten 1981". *Statistische Studieën 1989, Nationaal Instituut voor de Statistiek*.

Levinson, D.M. (2012). "Accessibility impacts of high-speed rail". *Journal of Transport Geography*, 22: 288–291.

Lhomme, S. (2017). "Développement d'un algorithme de détection de communautés spatiales". In *Spatial Analysis and GEOmatics 2017*. INSA de rouen, Rouen, France. URL https://hal.archives-ouvertes.fr/hal-01649121.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., and Ma, W.Y. (2008). "Mining user similarity based on location history". p. 34. ACM.

Li, X. and Farber, S. (2016). "Spatial representation in the social interaction potential metric: an analysis of scale and parameter sensitivity". *Journal of Geographical Systems*, 18(4): 331–357. URL https://link.springer.com/article/10.1007/s10109-016-0238-6.

Liu, Y., Sui, Z., Kang, C., and Gao, Y. (2014). "Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data". *PLOS ONE*, 9(1): e86 026. URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0086026.

Lombard, J. (1999). "Territoires, lieux et liens. Relations et savoir-faire de transporteurs de Saint-Omer (France)". *Les Cahiers Scientifiques du Transport*, (36).

Longley, P.A., Adnan, M., and Lansley, G. (2015). "The geotemporal demographics of Twitter usage". *Environment and Planning A abstract*, 47(2): 465–484. URL http://www.envplan.com/epa/fulltext/a47/a130122p.pdf.

Longley, P.A., Goodchild, M., Maguire, D.J., and Rhind, D.W. (2010). *Geographic Information Systems and Science*. Wiley Publishing.

Luyten, S. and Van Hecke, E. (2007). "De belgische stadsgewesten 2001". Technical report, Instituut voor Sociale en Economische Geografie, K.U.Leuven. URL http://statbel.fgov.be/nl/binaries/p009n014_nl%5B1%5D_tcm325-34289.pdf.

Ma, X., McCormack, E., and Wang, Y. (2011). "Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program". *Transportation Research Record: Journal of the Transportation Research Board*, 2246: 92–100. URL https://trrjournalonline.trb.org/doi/abs/10.3141/2246-12.

Macharis, C., Dobruszkes, F., and Hubert, M. (2014). "Mobilité et logistique à Bruxelles". Technical Report 2013/154341, ULB – Universite Libre de Bruxelles. URL https://ideas.repec.org/p/ulb/ulbeco/2013-154341.html.

Macharis, C. and Melo, S. (2011). *City distribution and urban freight transport: multiple perspectives*. Edward Elgar Publishing.

Marissal, P., Medina Lockhart, P., Vandermotten, C., and Van Hamme, G. (2006). "Les structures socio-économiques de l'espace belge". *Monographie N*, 6.

Mathä, T. and Wintr, L. (2009). "Commuting flows across bordering regions: a note". *Applied Economics Letters*, 16(7): 735–738.

Merenne-Schoumaker, B., Vandermotten, C., Van Hecke, E., Decroly, J.M., Vanneste, D., and Verhetsel, A. (2015). *Atlas de Belgique. Tome 5: Activités économiques*. Academia Press.

Miller, H.J. (2017). "Geographic information science II: Mesogeography: Social physics, GIScience and the quest for geographic knowledge". *Progress in Human Geography*, p. 0309132517712154. URL http://dx.doi.org/10.1177/0309132517712154.

Miller, H.J. and Goodchild, M.F. (2015). "Data-driven geography". *GeoJournal*, 80(4): 449–461. URL https://doi.org/10.1007/s10708-014-9602-6.

Moniteur Belge (2004). *Ordonnance portant assentiment à la Convention du 4 avril 2003 entre l'Etat fédéral, la Région flamande, la Région wallone et la Région de Bruxelles-Capitale, visant à mettre en oeuvre le programme du réseau express r'egional de, vers, dans et autour de Bruxelles*. Published: Moniteur Belge.

Montis, A.D., Caschili, S., and Chessa, A. (2013). "Commuter networks and community detection: A method for planning sub regional areas". *The European Physical Journal Special Topics*, 215(1): 75–91. URL https://link.springer.com/article/10.1140/epjst/e2013-01716-4.

Newman, M.E.J. (2003). "Mixing patterns in networks". *Physical Review E*, 67(2): 026 126. URL https://link.aps.org/doi/10.1103/PhysRevE.67.026126.

Newman, M.E.J. (2004). "Detecting community structure in networks". *The European Physical Journal B - Condensed Matter*, 38(2): 321–330. URL http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1140/epjb/e2004-00124-y.

Newman, M.E.J. and Girvan, M. (2004). "Finding and evaluating community structure in networks". *Physical Review E*, 69(2): 026 113. URL http://link.aps.org/doi/10.1103/PhysRevE.69.026113.

Nijkamp, P., Rietveld, P., and Voogd, H. (1990). *Multicriteria evaluation in physical planning*. North-Holland Amsterdam.

Observatoire de la Santé et du Social Bruxelles (2006). "Atlas de la santé et du social de Bruxelles Capitale". Technical report, Observatoire de la santé et du social Bruxelles.

O'Kelly, M.E. (2015). "Spatial Interaction". In J.D. Wright, editor, *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pp. 152–156. Elsevier, Oxford. URL http://www.sciencedirect.com/science/article/pii/B9780080970868720629.

O'Kelly, M.E., Niedzielski, M.A., and Gleeson, J. (2012). "Spatial interaction models from Irish commuting data: variations in trip length by occupation and gender". *Journal of Geographical Systems*, 14(4): 357–387. URL https://link.springer.com/article/10.1007/s10109-011-0159-3.

Onnela, J.P., Arbesman, S., González, M.C., Barabási, A.L., and Christakis, N.A. (2011). "Geographic Constraints on Social Network Groups". *PLoS ONE*, 6(4): e16 939. URL http://dx.plos.org/10.1371/journal.pone.0016939.

Papadopoulos, S., Kompatsiaris, Y., Vakali, A., and Spyridonos, P. (2011). "Community detection in Social Media". *Data Mining and Knowledge Discovery*, 24(3): 515–554. URL http://link.springer.com/article/10.1007/s10618-011-0224-z.

Pluvinet, P., Gonzalez-Feliu, J., and Ambrosini, C. (2012). "GPS Data Analysis for Understanding Urban Goods Movement". *Procedia - Social and Behavioral Sciences*, 39: 450–462. URL http://www.sciencedirect.com/science/article/pii/S1877042812005885.

Poelmans, L. and Van Rompaey, A. (2009). "Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region". *Landscape and Urban Planning*, 93(1): 10–19. URL http://linkinghub.elsevier.com/retrieve/pii/S0169204609001078.

Pucci, P., Manfredini, F., and Tagliolato, P. (2015). "Mobility Practices and Mobile Phone Data". In P. Pucci, F. Manfredini, and P. Tagliolato, editors, *Mapping Urban Practices Through Mobile Phone Data*, SpringerBriefs in Applied Sciences and Technology, pp. 1–11. Springer International Publishing, Cham. URL https://doi.org/10.1007/978-3-319-14833-5_1.

Pumain, D. (2003). "Une approche de la complexité en géographie". *Géocarrefour*, 78(Vol. 78/1): 25–31. URL http://journals.openedition.org/geocarrefour/75.

Pumain, D. (2017). "Geography Far from Equilibrium". In J. Johnson, A. Nowak, P. Ormerod, B. Rosewell, and Y.C. Zhang, editors, *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking*, Understanding Complex Systems, pp. 71–80. Springer International Publishing, Cham. URL https://doi.org/10.1007/978-3-319-42424-8_5.

Puura, A., Silm, S., and Ahas, R. (2018). "The Relationship between Social Networks and Spatial Mobility: A Mobile-Phone-Based Study in Estonia". *Journal of Urban Technology*, 25(2): 7–25. URL https://doi.org/10.1080/10630732.2017.1406253.

Quesnot, T. (2016). "L'involution géographique : des données géosociales aux algorithmes". *Netcom. Réseaux, communication et territoires*, (30-3/4): 281–304. URL http://journals.openedition.org/netcom/2545.

Rae, A. and Singleton, A. (2015). "Putting big data in its place: a Regional Studies and Regional Science perspective". *Regional Studies, Regional Science*, 2(1): 1–5. URL http://www.tandfonline.com/doi/abs/10.1080/21681376.2014.990678.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., and Strogatz, S.H. (2010). "Redrawing

the Map of Great Britain from a Network of Human Interactions". *PLoS ONE*, 5(12): e14 248. URL http://dx.plos.org/10.1371/journal.pone.0014248.

Reades, J., Calabrese, F., Sevtsuk, A., and Ratti, C. (2007). "Cellular census: Explorations in urban data collection". *Pervasive Computing, IEEE*, 6(3): 30–38. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4287441.

Reichardt, J. and Bornholdt, S. (2004). "Detecting Fuzzy Community Structures in Complex Networks with a Potts Model". *Physical Review Letters*, 93(21): 218 701. URL https://link.aps.org/doi/10.1103/PhysRevLett.93.218701.

Riguelle, F., Thomas, I., and Verhetsel, A. (2007). "Measuring urban polycentrism: a European case study and its implications". *Journal of Economic Geography*, 7(2): 193–215. URL http://joeg.oxfordjournals.org/content/7/2/193.abstract.

Rodrigue, J.P., Comtois, C., and Slack, B. (2013). *The geography of transport systems*. Routledge, London.

Rodrigue, J.P., Comtois, C., and Slack, B. (2016). *The geography of transport systems*. Routledge.

Rokach, L. and Maimon, O. (2005). "Clustering Methods". In O. Maimon and L. Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pp. 321–352. Springer US, Boston, MA. URL https://doi.org/10.1007/0-387-25465-X_15.

Rosvall, M. and Bergstrom, C.T. (2008). "Maps of random walks on complex networks reveal community structure". *Proceedings of the National Academy of Sciences*, 105(4): 1118–1123. URL http://www.pnas.org/content/105/4/1118.short.

Roy, J.R. and Thill, J.C. (2004). "Spatial interaction modelling". *Papers in Regional Science*, 83(1): 339–361. URL https://onlinelibrary.wiley.com/doi/abs/10.1007/s10110-003-0189-4.

Rozenblat, C. and Melancon, G., editors (2013). *Methods for Multilevel Analysis and Visualisation of Geographical Networks*. Methodos Series. Springer Netherlands. URL https://www.springer.com/gp/book/9789400766761.

RTBF (2016). "Une regionalisation de la SNCB est-elle possible?" URL https://www.rtbf.be/info/belgique/detail_une-regionalisation-de-la-sncb-est-elle-realisable?id=9178419.

Sagl, G., Delmelle, E., and Delmelle, E. (2014). "Mapping collective human activity in an urban environment based on mobile phone data". *Cartography and Geographic Information Science*, 41(3): 272–285. URL http://www.tandfonline.com/doi/abs/10.1080/15230406.2014.888958.

Sakarovitch, B., Bellefon, M.P.d., Givord, P., and Vanhoof, M. (2018). "Estimating the Residential Population from Mobile Phone Data, an Initial Exploration". *Economie et Statistique*, 505(1): 109–132. URL https://www.persee.fr/doc/estat_0336-1454_2018_num_505_1_10870.

Sapena, M., Ruiz, L.A., and Goerlich, F.J. (2016). "Analysing relationships between urban land use fragmentation metrics and socio-economic variables." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.

Scholtus, S. (2015). "Aantekeningen over het toewijzingsalgoritme voor Daytime Population".

Scott, A.J. and Storper, M. (2015). "The Nature of Cities: The Scope and Limits of Urban Theory: The nature of cities". *International Journal of Urban and Regional Research*, 39(1): 1–15. URL http://doi.wiley.com/10.1111/1468-2427.12134.

Servais, M., Thomas, I., Van Dyck, H., and Verhetsel, A. (2004). "Polycentrisme urbain : une réalité spatialement mesurable ?" *Cybergeo : European Journal of Geography*. URL http://journals.openedition.org/cybergeo/3669.

Sevtsuk, A. and Ratti, C. (2010). "Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks". *Journal of Urban Technology*, 17(1): 41–60. URL http://dx.doi.org/10.1080/10630731003597322.

Shen, L. and Stopher, P.R. (2014). "Review of GPS Travel Survey and GPS Data-Processing Methods". *Transport Reviews*, 34(3): 316–334. URL https://doi.org/10.1080/01441647.2014.903530.

Shen, Y. and Batty, M. (2018). "Delineating the perceived functional regions of London from commuting flows". *Environment and Planning A: Economy and Space*, p. 0308518X18786253. URL https://doi.org/10.1177/0308518X18786253.

SNCB (2014). *Nombre de voyageurs montes par gare en 2014*. URL https://www.belgianrail.be/fr//media/73AECBB3473141C1AA565C122AB6259C.ashx.

SNCB (2015). *Nombre de voyageurs montes par gare en 2015*. URL http://www.belgianrail.be/fr//media/8F764D77F60F48B188A8742652C6E48F.ashx.

Sohn, J. (2005). "Are commuting patterns a good indicator of urban spatial structure?" *Journal of Transport geography*, 13(4): 306–317.

Song, C., Qu, Z., Blumm, N., and Barabási, A.L. (2010). "Limits of Predictability in Human Mobility". *Science*, 327(5968): 1018–1021. URL https://science.sciencemag.org/content/327/5968/1018.

Spiekermann, K. and Wegener, M. (1994). "The shrinking continent: new time—space maps of europe". *Environment and Planning B: Planning and Design*, 21(6): 653–673.

Spiekermann, K. and Wegener, M. (1996). "Trans-European networks and unequal accessibility in Europe". *European journal of regional development*, 4(96): 35–42.

Statbel (2019). "Secteurs statistiques | Statbel". URL http://statbel.fgov.be/fr/propos-de-statbel/methodologie/classifications/secteurs-statistiques.

Steenbruggen, J., Tranos, E., and Nijkamp, P. (2015). "Data from mobile phone operators: A tool for smarter cities?" *Telecommunications Policy*, 39(3): 335–346. URL http://www.sciencedirect.com/science/article/pii/S0308596114000603.

Steinhaeuser, K. and Chawla, N.V. (2010). "Identifying and evaluating community structure in complex networks". *Pattern Recognition Letters*, 31(5): 413–421. URL http://linkinghub.elsevier.com/retrieve/pii/S0167865509003043.

Strale, M. (2009). "La mise en place d'une politique wallonne de promotion des activités logistiques; quels enjeux pour le territoire régional?" *Territoire (s) Wallon (s)*, (3): 191–202.

Strale, M. (2017). "Quelle place pour les activités portuaires et logistiques à Bruxelles ?" *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles / Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels*. URL http://brussels.revues.org/1491.

Strehl, A. and Ghosh, J. (2002). "Cluster ensembles—a knowledge reuse framework for combining multiple partitions". *Journal of machine learning research*, 3(Dec): 583–617. URL http://www.jmlr.org/papers/v3/strehl02a.html.

Thakur, A., Pinjari, A.R., Zanjani, A.B., Short, J., Mysore, V., and Tabatabaee, S.F. (2015). "Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips". *Transportation Research Record: Journal of the Transportation Research Board*, 2529: 66–73. URL https://trrjournalonline.trb.org/doi/abs/10.3141/2529-07.

Thisse, J.F. and Thomas, I. (2010a). "Brussels within the Belgian Economy : a geo-economic approach". Technical report. URL http://rethinkingbelgium.eu/rebel-initiative-files/ebooks/ebook-7/Re-Bel-e-book-7.pdf.

Thisse, J.F. and Thomas, I. (2010b). "Bruxelles au sein de l'économie belge : un bilan". *Regards économiques*. URL http://www.regards-economiques.be/index.php?option=com_reco&view=article&cid=91.

Thomas, I., Adam, A., and Verhetsel, A. (2017). "Migration and commuting interactions fields: a new geography with community detection algorithm?" *Belgeo. Revue belge de géographie*. URL http://journals.openedition.org/belgeo/20507.

Thomas, I., Cotteels, C., Jones, J., and Peeters, D. (2012). "Revisiting the extension of the Brussels urban agglomeration: new methods, new data... new results?" *Belgeo. Revue belge de géographie*. URL http://belgeo.revues.org/6074.

Thomas, I., Frankhauser, P., and De Keersmaecker, M.L. (2007). "Fractal dimension versus density of built-up surfaces in the periphery of Brussels*". *Papers in Regional Science*, 86(2): 287–308. URL http://dx.doi.org/10.1111/j.1435-5957.2007.00122.x.

Thomas, I., Hermia, J.P., Vanelslander, T., and Verhetsel, A. (2003). "Accessibility to freight transport networks in Belgium: a geographical approach". *Tijdschrift voor economische en sociale geografie*, 94(4): 424–438. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9663.00271.

Thomas, I., Jones, J., Caruso, G., and Gerber, P. (2018). "City delineation in European applications of LUTI models: review and tests". *Transport Reviews*, 38(1): 6–32. URL https://doi.org/10.1080/01441647.2017.1295112.

Thomas, I., Verhetsel, A., and Lorant, V. (2009). "Le recensement de la population : un requiem ?" Technical Report 67, Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES). URL https://ideas.repec.org/p/ctl/louvrg/67.html.

Thomopoulos, N., Givoni, M., and Rietveld, P. (2015). *ICT for Transport: Opportunities and Threats*. Edward Elgar Publishing.

Théré, C. and Séguy, I. (1998). "Huit siècles de découpages administratifs". Technical report, INSEE, Paris.

Timms, D.W.G. (1970). "Comparative factorial ecology: some New Zealand examples". *Environment and Planning A*, 2(4): 455–467.

Tobler, W.R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". *Economic Geography*, 46(sup1): 234–240. URL https://www.tandfonline.com/doi/abs/10.2307/143141.

Trasarti, R., Olteanu-Raimond, A.M., Nanni, M., Couronné, T., Furletti, B., Giannotti, F., Smoreda, Z., and Ziemlicki, C. (2015). "Discovering urban and country dynamics from mobile phone data with spatial correlation patterns". *Telecommunications Policy*, 39(3): 347–362. URL http://www.sciencedirect.com/science/article/pii/S0308596113002012.

Van Criekingen, M. (2002). "Les impacts sociaux de la rénovation urbaine à Bruxelles : analyse des migrations intra-urbaines". *Belgeo. Revue belge de géographie*, (4): 355–376. URL http://belgeo.revues.org/16092.

Van Criekingen, M. (2006). "Que deviennent les quartiers centraux à Bruxelles ?" *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles / Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels*. URL http://brussels.revues.org/293.

Van der Haegen, H. and Pattyn, M. (1980). "An Operationalization of the Concept of City Region in West-European Perspective: The Belgian City Regions". *Tijdschrift voor economische en sociale geografie*, 71(2): 70–77.

Van Der Herten, B., Van Meerten, M., and Verbeurgt, G. (2001). *Sporen in Belgie: 175 jaar spoorwegen, 75 jaar NMBS*. Universitaire Pers Leuven.

Van der Haegen, H., Van Hecke, E., and Savenberg, S. (2000). "Belgians on the move. Population distribution from a historical and modern perspective". *Belgeo. Revue belge de géographie*, (1-2-3-4): 173–188. URL http://journals.openedition.org/belgeo/14073.

Van Hamme, G., Grippa, T., and Van Criekingen, M. (2016). "Mouvements migratoires et dynamiques des quartiers à Bruxelles". *Brussels Studies. La revue scientifique électronique pour les recherches sur Bruxelles*

/ Het elektronisch wetenschappelijk tijdschrift voor onderzoek over Brussel / The e-journal for academic research on Brussels. URL http://journals.openedition.org/brussels/1331.

Van Hamme, G., Wertz, I., and Biot, V. (2011). "La croissance économique sans le progrès social: l'état des lieux à Bruxelles". *Brussels Studies*, 48: 1–20. URL http://www.briobrussel.be/assets/andere%20publicaties/fr_135_brus48fr.pdf.

Van Hecke, E., Halleux, J.M., Decroly, J.M., and Merenne-Schoumaker, B. (2009). "Noyaux d'habitat et Régions urbaines dans une Belgique urbanisée". Technical report, Direction générale Statistique et Information économique. URL http://orbi.ulg.be/handle/2268/66527.

Vance, J.E. (1964). *Geography and urban evolution in the San Francisco Bay area*. Institute of Governmental Studies, University of California.

Vandenbulcke, G., Dujardin, C., Thomas, I., de Geus, B., Degraeuwe, B., Meeusen, R., and Panis, L.I. (2011). "Cycle commuting in Belgium: spatial determinants and 're-cycling'strategies". *Transportation Research Part A: Policy and Practice*, 45(2): 118–137.

Vandermotten, C. (2016). "Adéquations et inadéquations du découpage territorial aux contextes politiques et économiques . Le cas de la Belgique". *EchoGéo*, (35). URL http://journals.openedition.org/echogeo/14490.

Vandermotten, C., Leclercq, E., Cassiers, T., and Wayens, B. (2009a). "L'économie bruxelloise". *Brussels Studies*, pp. 1–13. URL http://www.lasso.be/sites/default/files/etatsgeneraux_note_economie_26-01-09.pdf.

Vandermotten, C., Leclercq, E., Cassiers, T., Wayens, B., and others (2009b). "Etats Généraux de Bruxelles, note de synthèse n 7. L'économie bruxelloise". *Brussels Studies*, pp. 1–13. URL http://dial.uclouvain.be/downloader/downloader.php?pid=boreal:150350&datastream=PDF_01.

Vanderstraeten, L. and Van Hecke, E. (2019). "Les régions urbaines en Belgique". *Belgeo. Revue belge de géographie*. URL http://journals.openedition.org/belgeo/32246.

Vanneste, D., Thomas, I., and Vanderstraeten, L. (2008). "The spatial structure(s) of the Belgian housing stock". *Journal of Housing and the Built Environment*, 23(3): 173–198. URL https://doi.org/10.1007/s10901-008-9111-3.

Verhetsel, A., Beckers, J., and De Meyere, M. (2018). "Assessing Daily Urban Systems: A Heterogeneous Commuting Network Approach". *Networks and Spatial Economics*. URL https://doi.org/10.1007/s11067-018-9425-y.

Verhetsel, A., Kessels, R., Goos, P., Zijlstra, T., Blomme, N., and Cant, J. (2015). "Location of logistics companies: a stated preference study to disentangle the impact of accessibility". *Journal of Transport Geography*, 42: 110–121. URL http://www.sciencedirect.com/science/article/pii/S0966692314002464.

Verhetsel, A., Thomas, I., and Beelen, M. (2010). "Commuting in Belgian metropolitan areas". *Journal of Transport and Land Use*, 2(3/4): 109–131. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.9227&rep=rep1&type=pdf.

Verhetsel, A., Van Hecke, E., and Thomas, I. (2009). "Le mouvement pendulaire en Belgique : les déplacements domicile-lieu de travail, les déplacements domicile-école". URL https://dial.uclouvain.be/pr/boreal/object/boreal:127252.

Vickerman, R. (1997). "High-speed rail in Europe: experience and issues for future development". *The annals of regional science*, 31(1): 21–38.

Villanueva, D. and Feijóo, A. (2018). "Comparison of logistic functions for modeling wind turbine power curves". *Electric Power Systems Research*, 155: 281–288. URL http://www.sciencedirect.com/science/article/pii/S0378779617304340.

Wang, F. (2000). "Modeling commuting patterns in Chicago in a GIS environment: A job accessibility perspective". *The Professional Geographer*, 52(1): 120–133.

Wang, F. (2001). "Explaining intraurban variations of commuting by job proximity and workers' characteristics". *Environment and Planning B: Planning and Design*, 28(2): 169–182.

Ward, J. (1963). "Hierarchical Grouping to Optimize an Objective Function". *Journal of the American Statistical Association*, 58(301): 236–244. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845.

Wilson, A.G. (1974). *Urban and regional models in geography and planning*. URL https://trid.trb.org/view/1172294.

Yuan, Y. and Raubal, M. (2012). "Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data". In N. Xiao, M.P. Kwan, M.F. Goodchild, and S. Shekhar, editors, *Geographic Information Science*, Lecture Notes in Computer Science, pp. 354–367. Springer Berlin Heidelberg.

Zandbergen, P.A. (2009). "Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning". *Transactions in GIS*, 13: 5–25. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9671.2009.01152.x/abstract.

Zanjani, A.B., Pinjari, A.R., Kamali, M., Thakur, A., Short, J., Mysore, V., and Tabatabaee, S.F. (2015). "Estimation of Statewide Origin–Destination Truck Flows From Large Streams of GPS Data". *Transportation Research Record: Journal of the Transportation Research Board*, 2494: 87–96. URL http://trrjournalonline.trb.org/doi/abs/10.3141/2494-10.

Zhong, C., Arisona, S.M., Huang, X., Batty, M., and Schmitt, G. (2014). "Detecting the dynamics of urban structure through spatial network analysis". *International Journal of Geographical Information Science*, 28(11): 2178–2199. URL https://doi.org/10.1080/13658816.2014.914521.

List of Figures

1.1 The Belgian urban regions as developed by Vanderstraeten and Van Hecke (2019). 6
1.2 Municipalities in the BCR. 7
1.3 A general organisation of the BCR. 8
1.4 Schematic outline of the thesis. 18

2.1 Partitions detected using different values of ρ, in multi-scale modularity (A, B, C, D communities). 26
2.2 Compactness of communities. 26
2.3 Proportion of in-commuters working in the municipality of Brussels (A) and in the municipality of Leuven (B). 28
2.4 Weighted degree of the municipalities within the former province of Brabant (including self-loops). 30
2.5 Stability method: Variation of information and number of communities detected by varying ρ. 31
2.6 Partitions detected within the former province of Brabant for different value of ρ: A = 0.5; B = 0.7; C = 0.9; D = 1; E = 1.5; F = 2. 32
2.7 Number of times that each partition is detected for different values of ρ. 34
2.8 Number of communities detected, internal and external weighted edges for each partition. 35
2.9 Partitions detected in the former province of Brabant (A) with and (B) without self-loops (eight communities). 38
2.10 Communities detected in Belgium and zoomed on the former province of Brabant (A), and detected in the former province of Brabant (B). 41
2.11 Communities detected by the Louvain Method (A) and by the Infomap algorithm (B), computed for Belgium but presented for the former province of Brabant. 42

2.12 Renumbering of the communities for five runs of the Louvain Method. 44
2.13 Partitions detected in Belgium (A) with and (B) without self-loops (10 communities). 45

3.1 Theoretical network of 44 entities. 51
3.2 Anabel method applied to the theoretical network in iteration 32 (high) and 42 (low). 52
3.3 Louvain Method applied to the theoretical network. 53
3.4 Comparison of the results for the methods Anabel and Louvain on an imaginary data set. 55
3.5 Residential basins based on the Anabel method – authorized inter cluster aggregation situation after 562 iterations. 57
3.6 A) Residential basins based on the Anabel method – authorized inter cluster aggregation after 290 iterations - B) Residential basins based on the Anabel method – authorized inter cluster aggregation after 536 iterations. 58
3.7 A) Communities detected using the Louvain Method (absolute date and $\rho = 1$) - B) Communities detected using the Louvain Method (relative data and $\rho = 1.25$). 59
3.8 Comparison of partitions from the Louvain Method and Anabel. 60

4.1 Phone basins based on cell phones billing addresses (Blondel *et al.*, 2010) (1) regional city, (2) major city , (3) provincial border. 69
4.2 Variation of Information (VI) versus ρ values for commuting movements between municipalities. 73
4.3 Proportion of internal and external edges and number of communities for different values of ρ for the 2011 commuting movements (at municipality level). 74
4.4 Optimal partition in communities when nodes are municipalities (A – $\rho = 0.9$) and statistical sectors (B – $\rho = 1.0$) - 2011 commuting data. 74
4.5 Square matrix of commuting movements appearing in Belgium. 76
4.6 Robustness of the partitions reported in Figure 4.4: % of times a node belongs to the same community in the 1,000 runs - 2011 commuting data. 77
4.7 Commuting communities for different values of ρ (A = 0,7 - B = 0,8 - C = 1,0 - D = 2,0). Hatched municipalities are those where the robustness is $< 90\%$. 78
4.8 Migration communities when $\rho = 1.0$. Hatched municipalities are those where the robustness is $< 90\%$. 79

4.9 Migration communities when $\rho = 3.0$ (A) and $\rho = 8.0$ (B). Hatched municipalities are those where the robustness is $< 90\%$. 80
4.10 Communities based on commutes to and from the workplace in 2011, on the scale of the statistical sectors for $\rho = 1$ (A) and municipalities for $\rho = 1$ (B). 82
4.11 Flow of commutes between communities 1 (Leuven), 2 (Brussels) and all of the other Belgian communities (a). 83
4.12 Principal commuting movements appearing in the former province of Brabant. 84
4.13 Square matrix of commuting movements appearing in the former province of Brabant. 85
4.14 Optimal division into communities ($\rho = 1$), based on residential migration data from 2011. 87

5.1 Railway network versus travel-demand network (cities and travel flows symbols proportional to their size/volume; the colour of each train station represents its community). 93
5.2 (a) Belgium administrative and urban structure, and (b) location of the train stations. 94
5.3 Rank - size distribution of the *iRail* requests by (a) origins and (b) destinations. 96
5.4 Communities in the *iRail* dataset for the (a) *General* and (b) *No Brussels* subsets (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set). 100
5.5 Communities detected in the *iRail* dataset during (a) *Weekdays* and (b) *Weekends* (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set). 103
5.6 Communities in the *iRail* dataset for the requests made in (a) *Dutch* and in (b) *French* (in legend: O = Other communities, with less than 5 nodes; N = Not present in the data set). . . . 104
5.7 Communities detected within the *RER* zone (note: O = Other communities, with less than 5 train stations). 106
5.8 Chorematic representation of the travel demand by train in Belgium. 108
5.9 robustness of the communities by subset of the *iRail* data set (number of times, over the 1,000 iterations that a node is affected to its final community, in %). 113

6.1 Transformation of raw GPS points into trips and Origin-Destination segments. 123
6.2 Number of journeys detected for different duration of stops. . . 125
6.3 Appearing and disappearing of the trucks' ID around 02:00. . . 126

6.4 Main steps of the detection and correction of inconsistencies, according to the occurrence of errors observed for each ID. . . . 127
6.5 Trucks in Belgium: average daily *traffic*. 130
6.6 Trucks in Belgium: daily averaged *connections*. 131
6.7 Absolute daily traffic (a) and relative daily traffic (b) of trucks having the Brussels agglomeration as Origin or Destination. . . 134
6.8 Absolute OD (a) and relative (b) of trucks having the Brussels agglomeration as Origin or Destination. 135
6.9 Proportion of GPS points emitted by countries of registration by day. 137
6.10 Daily evolution of the proportion of A) the number of trucks per hour and B) the number of GPS points emitted per hour, for the four defined categories of trucks. 138
6.11 Communities of places located in Belgium and visited by truck: A) general network, B) places visited more than 10 times, C) *intra* trucks, D) *inter* trucks, E) *Belgian* trucks, and, F) *foreign* trucks. 140
6.12 Communities of places located in Brabant and visited by truck: A) general network, B) places visited more than 10 times, C) *intra* trucks, D) *inter* trucks, E) *Belgian* trucks, and, F) *foreign* trucks. 141

7.1 Number of mobile phone calls per day (A) and per hour on one working day (B). 151
7.2 Location of the studied calls within the Former province of Brabant (in % of the total). 152
7.3 Mobile telephony areas on the scale of Thiessen polygons around antennas (A) and on the scale of municipalities (B). 153
7.4 Mobile phone calls: communities of antennas detected for the four predefined networks: *General* (A), *Day* (B), *Weekend* (C) and *Night* (D). 155
7.5 Communities of antennas: uncertainty of the classification of the day (A) and night (B) partitions. 156
7.6 Circular graph and the associated O-D matrix (expressed in %) for all phone calls (*General*). 157
7.7 Mobile telephony areas (close-up of the centre of Brussels) during the day (A) and night (B) where locations on the map are: 1. Marché matinal et centre Européen de fruits et de légumes, 2. Quai des péniches; 3. Place Rogier, 4. Université Saint-Louis - Bruxelles (USL-B), 5. VRT - RTBF, 6. Gare Bruxelles Congrès, 7. Gare Schuman, 8. Gare Bruxelles-Luxembourg, 9. Vrije Universiteit Brussel (VUB), 10. Audi Forest and other industrial areas, 11. Bois de la Cambre. 158

7.8 Temporal variation of phone activity: A: number of calls emitted by four randomly selected antenna, B: cumulated calls emitted by each antenna, for all antennas, C: average cumulated calls for each cluster of antennas. 159
7.9 Clusters of antennas based on logistic curve characteristics of their hourly workday's activity. 160
7.10 Composition of the five clusters of temporal curves: A: parameter b (steepness of the curve), B: parameter d (curve max value), C: parameter g (Sigmoid midpoint). 161
7.11 Zoom on Brussels (A) and Ottignies Louvain-la-Neuve (B) where locations on the map are: 1 the Ixelles cemetery, 2 is the Eugène Flagey Place, 3 the Grand Place, 4 VW-Audi car factory (Forest), 5 Brussels Airport (Zaventem), 6 and 7 Sport centres, 8 Concert Hall (Forest National), 9 Ottignies, 10 family residential area, 11 student residential area, 12 Place des Sciences (party place) and 13 Industrial zoning areas (Parc Einstein, Parc Fleming and Axis Parc). 162
7.12 Clusters of statistical sections based on socio-economic variables. 164
7.13 Composition of each cluster of the concluding partition mapped in Figure 7.14. 167
7.14 A concluding partition of the Metropolitan area of Brussels based on the three classifications. 168
7.15 Cluster composition: A: dpop, B: Unemploym, C: commut.in, D: lessthan15 y., E: builtbef.1946, F: rentedresid. 170
7.16 Theoretical logistic curve. 172
7.17 Creation of the new BSU based on two different spatial entities. 173

List of Tables

2.1 Weighted degree size-ranked municipalities within the former province of Brabant. 29
2.2 NMI values between partitions resulting from different ρ-values. 33
2.3 Compactness of the communities detected in the province of Brabant for different values of ρ. 36
2.4 *Compactness* values of the communities detected with the inclusion of or without self-loops (the number of municipalities within communities is given in brackets). 40

3.1 Number of entities composing each group (basin/community), according to the method used and the number of groups selected (2 or 3). 54
3.2 Composition of basins (Anabel) and communities (Louvain Method). 61
3.3 Index c comparing basins and communities. 62

4.1 NMI values between pairs of partitions (municipality level). . . 81
4.2 Comparing solutions for partitions based on municipalities. . . 81

5.1 Descriptive statistics of the subsets. 99
5.2 Average travel distance alongside the railway network (km; between bracket = standard deviation). 99
5.3 Average value of the spatial contiguity index (between bracket = standard deviation). 100
5.4 Robustness of the communities (in %; high values = high robustness; between bracket = standard deviation). 101
5.5 Central node of the communities by betweenness centrality (between bracket = name of the train station). 114

6.1 Sensitivity analyses of the temporal criterion 124

6.2 Retained information after the main filtering, cleaning and validation steps. 128

7.1 NMI measured between pairs of calls network. 155
7.2 Definition of the variables. 171
7.3 Pearson correlation coefficients between the variables. 172