A graph based approach for functional urban areas delineation

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

A graph based approach for functional urban areas delineation

Noudéhouénou L. J. Houssou
Laboratoty L3I
La Rochelle, France
noudehouenou.houssou1@univ-lr.fr

Jean-loup Guillaume
Laboratoty L3I
La Rochelle, France
jean-loup.guillaume@univ-lr.fr

Armelle Prigent
Laboratoty L3I
La Rochelle, France
armelle.prigent@univ-lr.fr

ABSTRACT

In an increasingly urbanized world, where cities are changing continuously, it is essential for policy makers to have access to regularly updated decision-making tools for an effective management of urban areas. An example of these tools is the delineation of cities into functional areas which provides knowledge on high spatial interaction zones and their socioeconomic composition. In this paper, we presented a method for the structural analysis of a city, specifically for the determination of its functional areas, based on communities detection in graph. The nodes of the graph correspond to geographical units resulting from a cartographic division of the city according to the road network. The edges are weighted using a Gaussian distance-decay function and the amount of spatial interactions between nodes. Our approach optimize the modularity to ensure that the functional areas detected have strong interactions within their borders but lower interactions outside. Moreover, it leverages on POIs' entropy to maintain a good socioeconomic heterogeneity in the detected areas. We conducted experiments using taxi trips and POIs datasets from the city of Porto, as a study case. Trough those experiments, we demonstrate the ability of our method to portray functional areas while including spatial and socioeconomic dynamics.

CCS CONCEPTS

• **Information systems** → **Geographic information systems**; **Data mining**; • **Applied computing** → **Sociology**; • **Mathematics of computing** → *Graph theory*; • **Human-centered computing** → *Geographic visualization*;

KEYWORDS

Functional areas, Community detection, Complex networks, Trajectories analysis

ACM Reference Format:
Noudéhouénou L. J. Houssou, Jean-loup Guillaume, and Armelle Prigent. 2019. A graph based approach for functional urban areas delineation. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19), April 8–12, 2019, Limassol, Cyprus.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3297280.3297341

SAC'19, April 8-12, 2019, Limassol, Cyprus

ACM ISBN 978-1-4503-5933-7/19/04…$15.00
https://doi.org/10.1145/3297280.3297341

1 INTRODUCTION

By 2050, cities will have 2.5 billion more inhabitants than today [1]. In other words, more than two-thirds of the world's population will be urban. This requires rethinking of urban management and planning methods to face the challenge. An approach resulting from this rethinking is to use information from city-dwellers activities to delineate cities, into functional areas which provide a better understanding of the spatial and socioeconomic composition of urban areas.

Identifying the functional urban areas is a hot topic in urban computing that has multiple applications like tourist circuits recommendation or the choice of new businesses locations, for example. However, it can be a tough task because, to the best of our knowledge, there is no harmonized definition of the notion of functional area and the cities' division obtained from proposed methods are difficult to validate [9].

The literature propose several methods for detecting functional areas including supervised learning methods [27, 29] and unsupervised ones [8, 13, 24, 26] which are more numerous. Among those unsupervised methods, mainly based on clustering, most use statistical approaches and others use graph-based approaches. The later, have the advantage to better encapsulate the interactions and thus are more suitable for us to identify the functional areas.

In this paper, we developed a graph-based approach to delineate functional areas combining trajectories data and POIs (Points of Interests) while the other graph-based approaches only rely on trajectories or mobility data.

Our methodology consists in (i) constructing a graph of formal regions using trajectories, (ii) identifying communities, i.e. connected groups of formal regions, using a community detection algorithm. The construction of the network is fundamental and we will show that parameters can influence the size and contiguity of functional areas, (iii) controlling the heterogeneity of the communities detected using POIs' entropy.

The remainder of the paper is organized as follows. First we give some definitions of the notion of functional area in Section 2. In Section 3 we present and discuss some previous works that are closely related to our proposal. Then, we describe our methodology to extract functional areas in Section 4 and we present some results on the city of Porto in Section 5. Finally we conclude in Section 6.

2 THE CONCEPT OF FUNCTIONAL AREA

There exist several definitions of the functional area concept in the literature. Even the designation slightly varies from functional area/region to functional urban area/region depending on the scale

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

(country or city) considered. We present here some of those definitions.

Antikainen [1], defines a functional urban area as a travel to work area that is an agglomeration of work places attracting the work force of surrounding area. Its most important quality is its ability to exceed administrative boundaries.

For Karlsson [14], a functional (urban) region is characterized by an agglomeration of activities and by an intra-regional transport infrastructure, facilitating a large mobility of people, products, and inputs within its interaction borders.

According to the OECD [20], a functional region is a territorial unit resulting from the organization of social and economic relations in that its boundaries do not reflect geographical particularities or historical events. It is thus a functional sub-division of territories. Usually, functional areas are organized around one or several nodes, with the surrounding areas linked to that (those) node(s) through different systems (transportation, communication, labour, trade) [6].

In [9], a functional region is described as a geographical region in which within-region interaction in terms of commuters travel to work flows is maximized, and between-region interaction is minimized.

Beyond these definitions which focus on socio-economic interactions, a functional area is also considered as a territory with a particular function (residential, commercial, tourist, etc.) that depends on human activities occurring on it. In other words, a functional area corresponds to a certain land use [11, 28].

From those definitions we can extract some fundamental principles concerning functional areas. First, they don't correspond to administrative delimitation and don't depend on geographical or historical characteristics. Secondly, they are structured around nodes and present stronger interactions within their boundaries than with the outside. Thirdly, they are heterogeneous as they gather many activities and/or places. Finally, they can be associated with specific land uses.

These specificities make the delineation of functional area a bottom-up process that require interaction (mobility, communication, trade or labor) data. In our case study we will rely on transportation data to identify functional areas.

3 RELATED WORK

A bunch of works related to functional urban areas delineation have been recently produced. Two main types of approaches emerge from it: the statistical ones and the graph-based ones. The articles describing statistical approaches, associate functional urban areas delineation with the identification of residential, business, administrative, touristic areas and so on. According to this conception, Yuan et al. [26], develop a method using DMR (Dirichlet Multinominal Regression) to identify functional regions in Beijing. They combine mobility patterns extracted from taxi trips and POIs data as entry of their model which is an improvement of LDA (Latent Dirichlet Allocation). Another LDA based method has been proposed by Gao et al. [11] to detect functional areas in ten US cities. However, this model just exploits POIs data. In [23], Tu et al. use remote sensing images combined with GSM positioning data to identify functional zones. They extract various features from both of the data sources then apply a hierarchical clustering to portray the functional areas. Other methods based on matrix factorization [24], EM (Expectation Maximization) algorithm [17], LRA (Low Rank Approximation) or K-means [28, 29] and using taxi trips data, social media data, bus smart card data, POIs data are also proposed in the literature.

However, studies on functional urban areas are not exclusively related to land use. Some papers present them as regions where socio-economic interactions are stronger within their borders than outside. We share this point of view because it corresponds to our perception of functional areas. The methods described in these papers are usually graph-based because graphs naturally model interactions. They also involve community detection which aims to find set of nodes more strongly inter-connected with each other than with the outside. For example, in [9], Farmer and Fotheringham formalize functional areas delineation as a community detection problem in a travel-to-work flows' network. Their method seeks to maximize the modularity in the community detection process using a spectral approach. Their study case was the whole Irish territory. Similarly, to discover urban social functional regions in Shanghai, Fan et al. [8] apply the Fast-Newman algorithm to detect communities in a Voronoï cells graph, built using taxi trajectories data. Then they identify hot functional areas from those communities and label them. For their part, Demsar et al [7] rely on an edge-based community detection method for the identification of functional areas in London. The particularity of those detected functional areas is that they overlap.

These graph-based methods which are more relevant for us, due to our conception of functional areas, only use mobility data and do not take into account the socio-economic composition of the territories. We think that it is a limitation since functional areas are not only regions of strong interactions, but also heterogeneous areas which gather different types of activities. Thus, by playing on the heterogeneity factor, it is possible to improve the quality of functional urban areas delineation process. This is why we propose a method, based on community detection and which enriches existing ones by combining the use of POIs with mobility data. We describe this new method in the next section.

4 METHODOLOGY

According to the definitions from Section 2, one of the main characteristics of functional areas is to have stronger interactions inside their borders than outside. Finding regions with such characteristics, has been formalized by [9], as a community detection problem that can be solved by modularity optimization.

Modularity is a measure to evaluate the quality of a partition of the vertices in a graph [19]. It is based on the idea that a community is a set of vertices more linked among them than to the outside (as can be a group of friends). It is formally defined by Equation 1 :

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

where A is the adjacency matrix of the graph, P is a partition of the vertices, C is a cluster of the partition P, i and j are two vertices of C with k_i, k_j their respective degree, m is the number of edges in the graph.

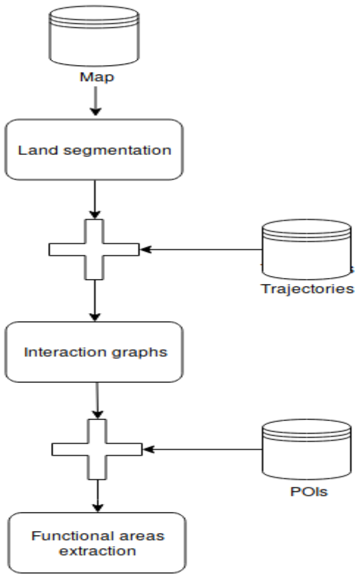

Figure 1: Workflow of functional areas detection using graph clustering.

The idea behind this formulation is to compare, for each community, the amount of internal links (sum of A_{ij}) with the expected number of internal links in a reference model (sum of $k_i k_j / 2m$). In this case the reference model is the configuration model [3] which generates a random graph that preserves the degrees of the original graph (all links are mixed but the degree of nodes is unchanged). Values taken by modularity are real numbers comprised between -0.5 and 1.

Our method to portray functional urban areas is inspired by [9] and also rely on modularity optimization but only partially. Indeed, unlike the authors, we integrate the points of interest information to consider the semantic and the heterogeneity of the processed areas. Moreover, we use a direct optimization algorithm as an alternative to the spectral method depicted in [9]. The three steps of our approach are: land segmentation, interaction graph construction and Functional areas extraction. Figure 1 illustrates the workflow of our method.

4.1 Land segmentation

Grid-based partitioning is a popular approach for dividing urban space [15, 16, 21, 25] because it is simple to implement and it evenly divide the study area. However, it does not integrate the natural urban layout, lacks semantic meaning [26] and is parameter dependent as it is necessary to fix the cells size.

To overcome these drawbacks, we subdivided the study area into disjoint spatial units using the road network. These spatial units can be assumed to have relatively homogeneous socio-economic functions. They are more suitable as land segmentation because they overcome the grid-based partitioning limitations. We will refer to these spatial units as formal regions as stated in [2].

4.2 Interaction graph construction

The interactions between formal regions are modeled by a graph of flows and represented as an Origin-Destination (OD) matrix. Each cell of the matrix indicates the number of trips performed between the corresponding two formal regions during a certain period. The matrix is symmetric as we ignore trips' directions. In order to obtain continuous zones, cell values are adjusted using a geographical weighting. The weight is computed with the Gaussian distance decay function presented in Equation 2 as proposed by [9]:

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

where A_{ij} is the weighted value of interaction between formal regions i and j, W_{ij} represent the number of trips between zones i and j, d_{ij} is the Euclidean distance between zones i and j, h is a threshold to control the bandwidth of the Gaussian function and also the compactness of the functional areas.

4.3 Functional areas extraction

Optimizing the modularity is an NP-complete problem [5]. That limitation leads to the development of heuristic approaches able to find good partitions faster and with fewer calculations than greedy approaches. According to [10], Louvain community detection algorithm [4] is one of the most efficient heuristic algorithms in the literature. Louvain performs repeatedly Two steps. A step of local optimization of the modularity, during which nodes are moved from community to community in order to increase the modularity, until a local maximum is reached. And a step that merges nodes inside communities to reduce the size of the network. It outperforms most of the other solutions both in computation time and quality of the modularity obtained by optimization. Reason why we choose it.

To extract functional areas, we apply Louvain algorithm to a range of interaction graphs obtained by varying the threshold value in the distance decay function. Then for each graph partition obtained we assess the heterogeneity of the corresponding regions by computing their average entropy, as defined in Equation 3 using the POIs distribution.

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

where E_R is the entropy defined by Shannon [22] of the region R and computed as:

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

with p_i the proportion of the i^{th} POIs' category in region R.

The intuition is that the more a region is heterogeneous the more its entropy is higher and the more there are reasons for spatial interactions inside it. Finally, we select the partition that maximize the product of modularity by average entropy as a multi-criteria selection rule to ensure that the functional areas generated have strong interactions as well as high heterogeneity.

5 RESULTS AND DISCUSSION

5.1 Datasets and study area

Our study area, Porto, is the core of the second largest urban area of Portugal, after Lisbon, with a population of 237591 inhabitants (census of 2011, http://mapas.ine.pt) and an area of 41.42 km^2. This ancient European center, also capital of the northern region of Portugal is bordered by the Douro river and the Atlantic Ocean.

Porto has an integrated transportation system combining buses, trams, metro and taxis. Some of the taxis are equipped with mobile data terminals able to provide various information on them. By collecting those information, the open dataset "Taxi Service Trajectory" has been built up and released as material for the ECML/PKDD 15 Kaggle competition. It describes trips performed by 442 taxis during an entire year from 01/07/2013 to 30/06/2014 in Porto [18]. Each trip record contains information about the trip ID, the type of request (dispatching from the central, direct demand to the driver on a stand, demand on a street), an anonymized identifier for the phone number used to make the demand, the origin stand of the taxi if the trip is asked directly on a stand, the taxi ID, the timestamp, the day type (holyday, workday or weekend), the lack or not of GPS data, a polyline containing a list of GPS coordinates (latitudes and longitudes) collected every 15 seconds. We used the aforementioned mobility dataset for the purposes of our research. The structure and some samples of the dataset are presented in Table 1.

The "Taxi Service Trajectory" dataset contains 1710670 records. After a slight preprocessing step to exclude the trips with missing data and those which are too short or too long we end up with 1438924 trajectories. The very short trajectories have their origin and destination associated with the same formal region while the very long may have origin inside the administrative border of Porto and destination outside and vice versa. We store the origins and destinations points in an indexed Postgresql database with the Postgis extension activated to manipulate and make calculations on spatial objects.

A total of 7710 points of interest have been collected from openstreetmap platform [2]. The POIs dataset contains various basic types (ATM, bank, church, hotels...) that we aggregated in 10 more generic groups as summarized in Table 2.

Finally, formal regions are extracted from the 2012 urban atlas available on the Portuguese Territorial General Directorate web site [3]. It is a vector shapefile with cadastral information on urban areas in Portugal including Porto. It does not indicate hierarchy level of roads but contains most of the roads observable in Porto. The land segmentation process results in 2453 formal regions. Figure 2 illustrates the obtained delineation of Porto into formal regions.

5.2 Experimentation and results

Our experiments are conducted on a HP Zbook laptop, containing an Intel i7-6700HQ octocore processor clocked at 8 * 2.60Ghz with 16Gb of Ram and running under Ubuntu 16.04 LTS. The GIS processing like the formal areas extraction from the urban atlas or the generation of functional areas maps are performed with QGIS.

[2]https://www.openstreetmap.org

[3]http://mapas.dgterritorio.pt/atom-dgt/CDG_atlasurbano2012_Continente_Atom.xml

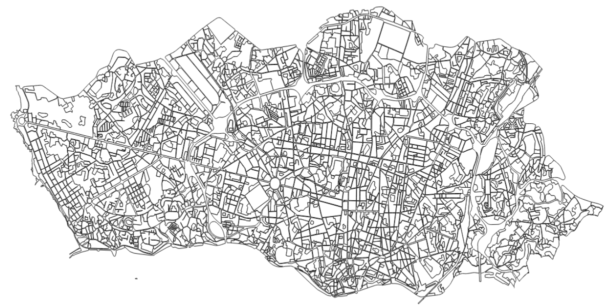

Figure 2: Porto formal regions.

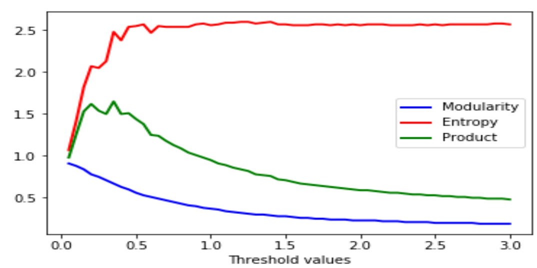

Figure 3: Modularity and mean entropy for different threshold values.

From formal regions and origin-destination data we built up a set of interaction graphs by varying the Gaussian threshold (parameter h in Equation 2) between 0.05 and 3 with a step of 0.05. Some of the formal regions don't have interactions associated and are not included in the interaction graphs construction.

Louvain algorithm is executed ten times on each interaction graph to deal with its non-determinism and the corresponding highest modularity partition is selected. Then we compute the average entropy of the detected areas. The results are presented in Figure 3.

The stop condition of the experimentation is the convergence of the average entropy. According to it, we choose 3 as upper bound of the threshold values given that the average entropy stagnates in the interval [1.5,3].

The best threshold value from Figure 3 is $h = 0.35$. This value maximizes the product of the average entropy and the modularity. The partition induced consists of eleven communities which correspond to functional areas as illustrated on Figure 4. Each community is identified by an ID and a color. We show in Figure 5 and 6 two examples of clustering using respectively $h = 0.1$ and $h = 2.0$. As stated before, the size of the clusters is directly impacted by the value of the threshold h even though we did not try to establish a direct correlation between these two parameters.

After analysis of the detected functional areas, the first remark is that those areas do not have the same boundaries as the administrative delineation of Porto into districts presented in Figure 7. This confirms that the city has a latent structure only accessible via the spatial interaction data. Activities diversity for each functional

Table 1: Dataset structure

Trip_id	Call_type	Origin_call	Origin_stand	Taxi_id	Timestamp	Day_type	Missing_Data	Polyline
1372637343620000571	A	31508	N/A	20000571	1372637343	A	False	[[8.618868,41.155101],...]
1372641742620000657	B	N/A	22	20000657	1372641742	A	False	[[8.689338,41.168124],...]
1372636858620000589	C	N/A	N/A	20000589	1372636858	A	False	[[8.618643,41.141412],...]

Table 2: POIs categorization

ID	Basics types	Categories
1	apartment, hotel, house, residential, dormitory, houseboat, guest_house, hostel, motel	Accomodation and Residence
2	courthouse, coworking_space, embassy, fire_station, police, post_office, accountant, company	Workplace and Public Service
3	bar, bbq, cafe, fast_food, ice_cream, pub, miniature_golf, nature_reserve, park, sports_centre, stadium	Sustenance and Leisure Places
4	aquarium, artwork, attraction, gallery, museum, theme_park, viewpoint	Tourism
5	train_station, bus_station, taxi, parking	Transportation
6	atm, bank, commercial, industrial, retail, warehouse, kiosk, boutique, fabric	Business and Industry
7	college, kindergarten, library, archive, school, music_school, language_school, university	Education
8	clinic, dentist, doctors, hospital, nursing_home, pharmacy, social_facility, veterinary, blood_donation	Healthcare
9	allotments, farmland, farmyard, forest, grass, greenfield, greenhouse_horticulture, meadow	Grass, Greenfield and Farmland
10	cemetery, church, chapel, place_of_worship	Worship

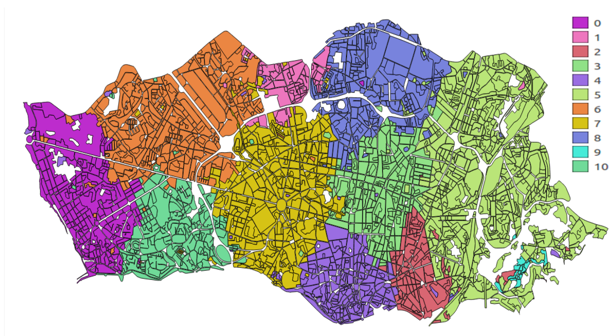

Figure 4: Functional areas obtained with a threshold $t = 0.35$.

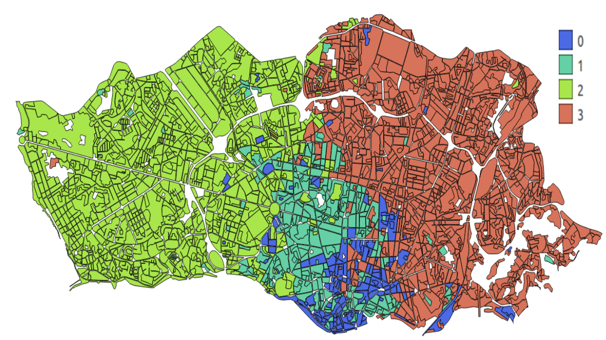

Figure 6: Functional areas obtained with a threshold $t = 2.0$.

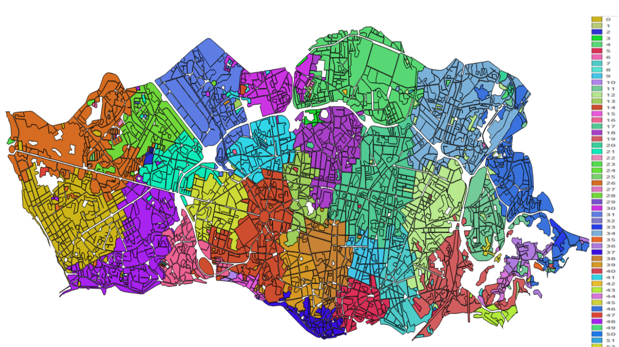

Figure 5: Functional areas obtained with a threshold $t = 0.1$.

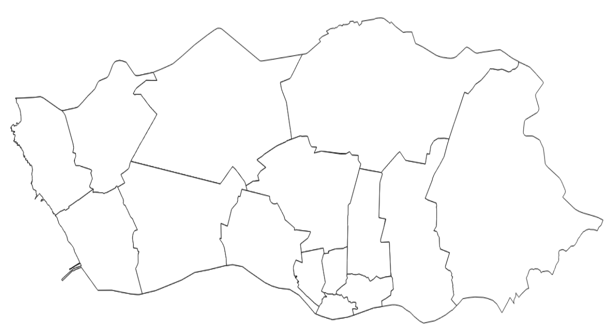

Figure 7: Districts of Porto.

area is presented in Figure 8 which illustrates the POIs distribution per community. Note that colors identifying detected functional areas and those associated with POIs distribution are not related.

From the Figure 8 we can observe that the communities detected have a good variety of POIs. Residential areas prevail in most of

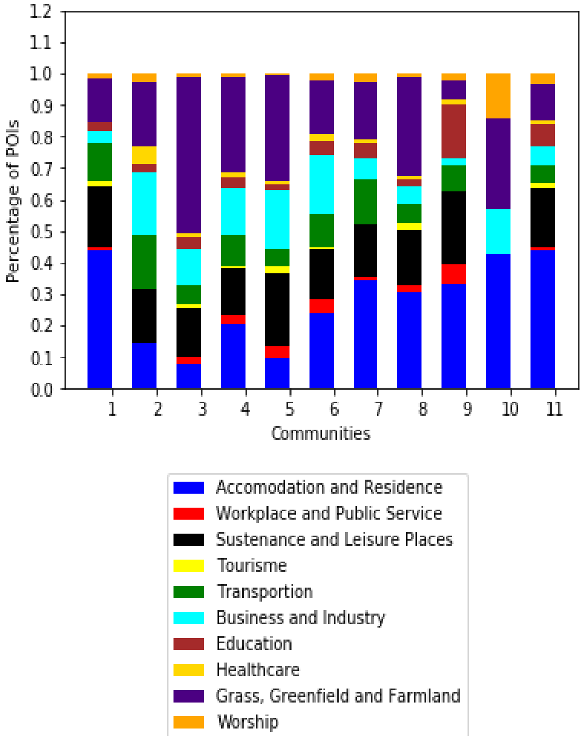

Figure 8: POIs distribution per cluster.

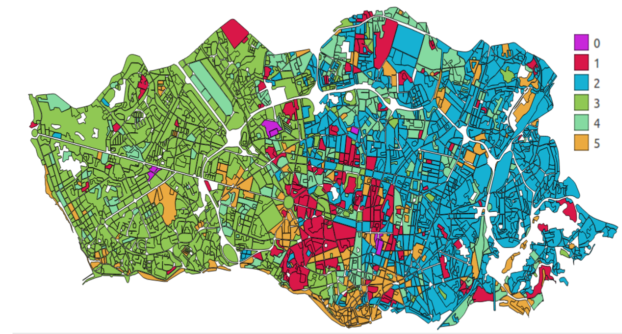

Figure 9: Functional areas with automatic threshold evaluation for each formal regions.

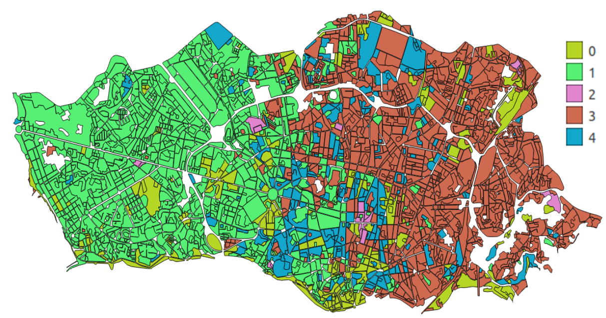

Figure 10: Functional areas with automatic threshold evaluation for all the formal regions.

the communities but this situation is normal since providing accommodation is one of the main, if not the main, function of a city.

Although the functional areas detected consist of large continuous blocks, it is sometimes observed that some of their formal regions are scattered outside their respective boundaries. It is because some formal regions are strongly interconnected despite their remoteness and conserve a significant interaction value in spite of the application of the distance decay function. To deal with this phenomenon we can either assign to the scattered formal regions the dominant clusters id of their neighborhood or keep them as is and use them as an indicator of the strong connections among functional areas.

According to the experimentation of [9], the bandwidth (h) for the distance weighting function should have been computed using an adaptive bandwidth selection procedure whereby a value of h is computed separately for each origin in the interaction matrix based on the distances to its corresponding destinations. However, this computation of the bandwidth result into very scattered functional areas as presented in Figure 9 when applied on our interaction graph, making it inefficient for the detection of functional areas at the level of a city. As a solution, we adopted a global bandwidth selection approach instead of the local one. It means that we determine a global bandwidth value for all the formal regions by fitting the distribution of distances with a Gaussian law. However, this bandwidth calculation also generates very scattered communities as portrayed in Figure 10. Due to the poor results obtained with these two automatic bandwidth selection methods, we were not able to propose a quantitative comparison of our method with the Farmer and Fotheringham approach which we used as baseline.

6 CONCLUSION AND FUTURE WORK

This paper proposes a graph-based method for the delineation of cities' functional areas, using communities detection. First, we divide the study area into disjoint spatial units called formal regions according to the road network. Secondly, we build weighted interaction graphs, using the formal regions and taxis trajectories. The nodes of the graph correspond to the formal regions and two nodes are connected if there exist taxis trips originating from one and ending in the other. The weights on the edges depend on the number of trips between the nodes and on the remoteness between them through a Gaussian distance decay function. We generate a set of interaction graphs by varying the threshold value of the Gaussian distance decay function. Thirdly, we apply to the obtained graphs the Louvain community detection algorithm in order to get partitions that maximize the modularity. Then, we compute the POIs' average entropy of detected partition for each interaction graph. Finally, we select the partition with the highest product of modularity by average entropy. Taxi trips and POIs dataset from Porto were used to experiment our method with coherent results.

This study can be extended to assess the stability of the algorithm and the detected areas in general and also in the case of time evolution. To that end, the dataset can be split in subsets with respect to time periods and functional areas detected for each subset. Then cores [12] will be computed over the different partitions obtained. In addition to that it would be useful, to formalize the relation between the size of functional areas and the Gaussian threshold values. Another perspective is the determination of overlapping communities to have a more realistic delineation of the city as one location can be strongly linked to other locations in disjoints functional areas. Moreover, we can leverage on the whole trajectory and the associated timestamps instead of the origin and destination points only to improve the method. For instance, using the whole trajectory, we can determine the real distance travelled on the road network and all the areas crossed to make a finer analysis. We could also compare functional urban areas detected using, separately or in combination, various kind of mobility data (taxi trajectories, bike-riding, buses trajectories, personal car trajectories). Lastly, the time dimension could help us to define a more accurate version of edge's weight.

REFERENCES

[1] J. Antikainen. 2005. The concept of functional urban area. Findings of the Espon project. *Informationen zur Raumentwicklung* 7 (2005), 447–452.

[2] S.W. Bednarz et al. 1994. *Geography for Life: National Geography Standards.* Distributed by ERIC Clearinghouse (Washington, D.C.). 278 p. pages. http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED375073

[3] E.A. Bender and E.R. Canfield. 1978. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A* 24, 3 (1978), 296 – 307. https://doi.org/10.1016/0097-3165(78)90059-6

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

[5] U. Brandes, D. Delling, M. Gaertler, R. Goerke, M. Hoefer, Z. Nikoloski, and D. Wagner. 2006. Maximizing Modularity is hard. http://arxiv.org/abs/physics/0608255 cite arxiv:physics/0608255 Comment: 10 pages, 1 figure.

[6] CSIL. 2015. *Territorial Agenda 2020 put in practice. Enhancing the efficiency and effectiveness of Cohesion Policy by a place-based approach.* Publications Office of the European Union. 53 pages. https://doi.org/10.2776/984537

[7] U. Demsar, J. Reades, E. Manley, and J.M. Batty. 2014. Edge-Based Communities for Identification of Functional Regions in a Taxi Flow Network. In *Extended Abstract Proceedings of the GIScience 2014*. 55–60.

[8] K. Fan, D. Zhang, Y. Wang, and S. Zhao. 2015. Discovering Urban Social Functional Regions Using Taxi Trajectories. In *Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing*. 356–359. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.77

[9] C.J.Q. Farmer and A.S. Fotheringham. 2011. Network-Based Functional Regions. *Environment and Planning A: Economy and Space* 43, 11 (2011), 2723–2741. https://doi.org/10.1068/a44136 arXiv:https://doi.org/10.1068/a44136

[10] S. Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3-5 (2010), 75 – 174. https://doi.org/DOI:10.1016/j.physrep.2009.11.002

[11] S. Gao, K. Janowicz, and H. Couclelis. 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS* 21, 3 (2017), 446–467. https://doi.org/10.1111/tgis.12289 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12289

[12] D. Gfeller, J.-C. Chappelier, and P. De Los Rios. 2005. Finding instabilities in the community structure of complex networks. *Phys. Rev. E* 72 (Nov 2005), 056135. Issue 5. https://doi.org/10.1103/PhysRevE.72.056135

[13] H. Han, X. Chen, and Y. Long. 2015. Discovering functional zones using bus smart card data and points of interest in Beijing. *CoRR* abs/1503.03131 (2015).

[14] C. Karlsson. 2007. *Clusters, Functional Regions and Cluster Policies.* Working Paper Series in Economics and Institutions of Innovation 84. Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies. https://ideas.repec.org/p/hhs/cesisp/0084.html

[15] X. Liu, L. Gong, Y. Gong, and Y. Liu. 2015. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography* 43 (2015), 78 – 90. https://doi.org/10.1016/j.jtrangeo.2015.01.016

[16] Y. Liu, F. Wang, Y. Xiao, and S. Gao. 2012. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning* 106, 1 (2012), 73 – 87. https://doi.org/10.1016/j.landurbplan.2012.02.012

[17] Y. Long and Z. Shen. 2015. *Discovering Functional Zones Using Bus Smart Card Data and Points of Interest in Beijing.* Springer International Publishing, Cham, 193–217. https://doi.org/10.1007/978-3-319-19342-7_10

[18] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas. 2013. Predicting Taxi-Passenger Demand Using Streaming Data. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (Sept 2013), 1393–1402. https://doi.org/10.1109/TITS.2013.2262376

[19] M.E.J. Newman. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 23 (2006), 8577–8582. https://doi.org/10.1073/pnas.0601602103 arXiv:http://www.pnas.org/content/103/23/8577.full.pdf

[20] OECD. 2002. *Redefining Territories: The Functional Regions.* Éditions OECD, Paris. 132 pages. https://doi.org/10.1787/9789264196179-en

[21] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou. 2014. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science* 28, 9 (2014), 1988–2007. https://doi.org/10.1080/13658816.2014.913794

[22] C.E. Shannon. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 3 (7 1948), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[23] W. Tu, Z. Hu, L. Li, J. Cao, J. Jiang, Q. Li, and Q. Li. 2018. Portraying Urban Functional Zones by Coupling Remote Sensing Imagery and Human Sensing Data. *Remote Sensing* 10, 1 (2018). https://doi.org/10.3390/rs10010141

[24] Y. Wang, Y. Gu, M. Dou, and M. Qiao. 2018. Using Spatial Semantics and Interactions to Identify Urban Functional Regions. *ISPRS International Journal of Geo-Information* 7, 4 (2018). https://doi.org/10.3390/ijgi7040130

[25] Y. Wang, T. Wang, M.-H. Tsou, H. Li, W. Jiang, and F. Guo. 2016. Mapping Dynamic Urban Land Use Patterns with Crowdsourced Geo-Tagged Social Media (Sina-Weibo) and Commercial Points of Interest Collections in Beijing, China. *Sustainability* 8, 11 (2016). https://doi.org/10.3390/su8111202

[26] J. Yuan, Y. Zheng, and X. Xie. 2012. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12).* ACM, New York, NY, USA, 186–194. https://doi.org/10.1145/2339530.2339561

[27] Z. Zhang, Y. Wang, Q. Liu, L. Li, and P. Wang. 2016. A CNN based functional zone classification method for aerial images. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).* 5449–5452. https://doi.org/10.1109/IGARSS.2016.7730419

[28] Y. Zhi, Y. Liu, S. Wang, M. Deng, J. Gao, and H. Li. 2014. Urban spatial-temporal activity structures: a New Approach to Inferring the Intra-urban Functional Regions via Social Media Check-In Data. *CoRR* abs/1412.7253 (2014).

[29] Z. Zhu, J. Yang, C. Zhong, J. Seiter, and G. Tröster. 2015. Learning Functional Compositions of Urban Spaces with Crowd-augmented Travel Survey Data. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '15).* ACM, New York, NY, USA, Article 22, 10 pages. https://doi.org/10.1145/2820783.2820832