AGILE: GIScience Series, 3, 63, 2022. https://doi.org/10.5194/agile-giss-3-63-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Mapping a historic neighbourhood through user-generated content: the case of Alfama, Lisbon (Portugal)

Vicente Tang [1], Jaskaran Puri [1], and Marco Painho [1]

[1]NOVA Information Management School, NOVA University Lisbon, Campus de Campolide, 1070-312 Lisbon

Correspondence: Vicente Tang (vtang@novaims.unl.pt)

Abstract. Participant-based methods aimed at extracting neighbourhood definitions are labour and time intensive. On the other hand, user-generated content (UGC) can provide locations to assess the extent of neighbourhoods. We investigated the definitions of Alfama - a historic neighbourhood in Lisbon (Portugal) - using six sources of UGC and applied a modification of the DBSCAN algorithm developed in the literature. By generating shapes from each source, we were able to visually and quantitatively evaluate their agreement as well as their differences. We demonstrate how different profiles of user activity from each source yielded varied geographies of Alfama. Although discrete representations are not the optimal choice, practical applications such as urban planning usually demand sharp definitions. Lastly, our approach can be extended and improved by adding more sources of UGC data and by picking other case studies.

Keywords. neighbourhoods, user-generated content, A-DBSCAN, Alfama

1 Introduction

Neighbourhoods are defined by one or a specific profile of different spatially-based attributes such as morphological, infrastructural, demographic, political and sentimental (Galster, 2001). They constitute an elementary social and spatial unit for urban geography, health sciences, social sciences, psychology and policy areas such as urban planning, education, resource allocation, emergency services, tourism management, crime, healthcare and others (Davies et al., 2009; Brindley et al., 2018; Catney et al., 2019). Dating back to the Middle Ages, Alfama is one of Lisbon's historic neighbourhoods and has been facing issues regarding gentrification, overtourism and touristification (Daly et al., 2021). From low-income and low-education levels in the early 2000's to having 35% of its real state currently directed to tourist accommodation, Alfama spreads across different official parishes in Lisbon (Sequera and Nofre, 2020; Madeira et al., 2021).

Alike other historic neighbourhoods, Alfama is not bounded to administrative borders and hence its extent may vary according to people's activities, perceptions and experiences (Twaroch et al., 2019). Although neighbourhoods are spatially vague, fetching their boundaries is important not only for research purposes but also for practical applications such as geographic information retrieval in location-based services as well as outlining units for planning and execution of public policies (Shang et al., 2016; Brindley et al., 2018). Nonetheless, collecting data directly from citizens, tourists and frequent visitors is a labour-intensive and lengthy process (Tang et al., 2021). Alternatively, the big geospatial data from user-generated content (UGC) has been extensively explored to harvest information on various dimensions of the urban (Goodchild, 2007; de Oliveira and Painho, 2021).

In this paper, we aim at generating and comparing the boundaries of the neighbourhood of Alfama obtained from six different sources of UGC. Although previous works have used simultaneously up to six web-based sources to extract city centres, neighbourhoods and cognitive regions, we introduce two new components to our analysis. First, we employ an extension of the density-based spatial clustering of applications with noise, the "Approximate DBSCAN" (A-DBSCAN), developed by Arribas-Bel et al. (2021). Aiming at retrieving sharp definitions of the neighbourhood extent, its main advantage consists in running iterations and computing different α-shapes, yielding optimal stable polygons given point-based distributions.

In addition, we split the data from social media (Twitter and Instagram) into portuguese and non-portuguese subsets to seek distinctions between the groups in the context of residents, tourism and visitors. Combined, the specific data sources, clustering method as well as the language subsets have not been implemented to outline the boundaries of a non-official historic neighbourhood. The Alfama case study is particularly worth exploring due to current issues on housing policies, cultural identity and conflicts between residents and tourists, although the workflow can be employed in any other city (Tulumello and Allegretti, 2021; Rêgo and Almeida, 2022).

2 Related work

Data on people's perceptions of neighbourhoods have been collected mostly through surveys or UGC. Online sources in previous works include websites, digital gazetteers, various social media platforms and points of interest (POI), with the objective of spatially assessing not only official administrative areas, but also other concepts such as *livehoods* (Cranshaw et al., 2012); vague neighbourhoods (Brindley et al., 2018); functional regions (Gao et al., 2017a) and areas of interest (AOI) (Hu et al., 2015; Mai et al., 2018). Although individuals have different mental maps, there is a significant degree of consensus regarding the spatial and semantic attributes of known neighbourhoods (Montello et al., 2003, 2014). As geo-tagged UGC connects human-generated information with spatial footprints, it acts as a proxy for in-space human-environment interaction (Papadakis et al., 2020). However, most sources of online data are bounded to point representations, with an assigned pair of geographic coordinates.

In the literature, methods aimed at extracting boundaries and extents of urban regions from point-based data vary. Previous studies have applied kernel density estimation (KDE) to generate surfaces, extracting regions using different thresholds (Jones et al., 2008; Hollenstein and Purves, 2010; Brindley et al., 2018; Twaroch et al., 2019). Other methods include the α-shape algorithms (Arampatzis et al., 2006); fuzzy logic approaches; (Schockaert et al., 2005); concave hulls (Hu et al., 2015, 2021); support-vector machines (Cunha and Martins, 2014); and Delaunay triangulation (Gao et al., 2017a). Prior to outlining boundaries and shapes, the DBSCAN algorithm has been widely used to identify prominent distributions of points. The A-DBSCAN - an extension proposed by Arribas-Bel et al. - performs replications using random data subsamples in order to find the most stable candidate solution. While the authors implemented the technique for delineating urbanized areas, the next section shows how we applied the method to discern the spatial extents of Alfama from different sources of UGC.

3 Methods

3.1 Data collection and pre-processing

Using official and non-official APIs, we retrieved data from Instagram, Twitter, Wikipedia, AirBnB, OpenStreetMap (OSM) and Idealista for the city of Lisbon and filtered the entries that contained the word "Alfama" in textual attributes. The latter source is a Portuguese rental listing and real-estate website widely used by residents.

We must acknowledge that the platforms differ in their geo-tagging procedures. While AirBnB and Idealista seem to have their listings with coordinates attached to buildings, Instagram does not track the precise location of their users, and instead their geo-tags are linked to points of interest.. On the other hand, Twitter allowed users to attach precise locations until 2019. From then on, precise coordinates can only be stored when users cross-post from Instagram or post pictures on the platform, whereas other geo-tags are linked to places and points of interest. These distinctions between sources can result in different levels of resolution, granularity and precision, which in turn might impact the results.

In addition, we manually downloaded the boundaries of the neighbourhood from OpenStreetMap as well as the shapes of the old parish distribution of the city. Although there are no official boundaries, the extent of Alfama used to be better represented by the combination of two former parishes: São Miguel and Santo Estêvão (Madeira et al., 2021).

For Twitter, Instagram and AirBnB, where few users might have a large number of entries, we attempted to reduce the bias by selecting a subsample from the more active users (Gao et al., 2017b). Based on the 90^{th} percentile of posts per user for each dataset (n), we randomly selected n posts for those users than contributed more than n times. We display in Fig. 1 the spatial distribution of point-based posts for each source after pre-processing (except OSM).

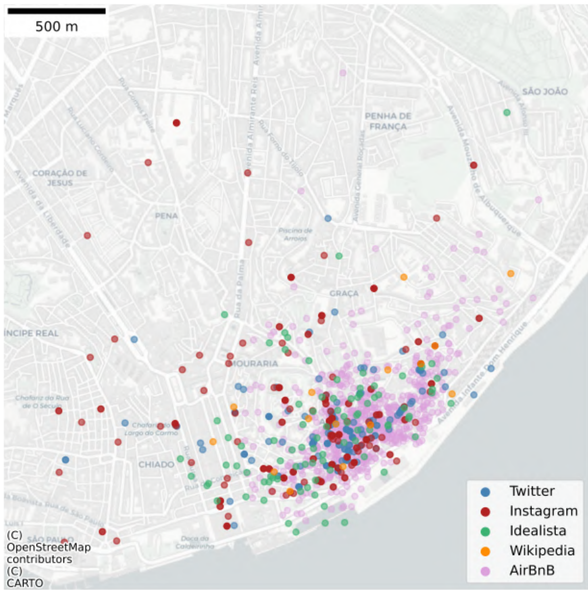

Figure 1. Distribution of selected posts from the different UGC sources across the city of Lisbon

3.2 Analysis

We first split the data coming from social media (Instagram and Twitter) into portuguese and non-portuguese language sub-datasets. By doing so, our goal was to create a proxy for online activity from foreigners and locals. Before running the A-DBSCAN algorithm, we ran a nearest neighbour (NN) analysis for each dataset and calculated their distributions' 90^{th} percentile values. We then obtained their average to be input as the *eps* parameter of the algorithm for all datasets.

As a single polygon, OSM data for Alfama did not go through the clustering analysis. In addition, because AirBnB data instances were unproportionally larger when compared to other sources (Fig. 1), its NN value was not included when computing the average value for *eps* to avoid spatial bias. The minimum sample size was the 3% of the count of samples for each dataset and we ran the method with 10 iterations (Gao et al., 2017b). The A-DBSCAN outputs were the optimal candidate solutions of clusters with their respective optimal α-shapes, which are generated to create tighter boundaries with less empty areas when compared to convex hulls (Chen et al., 2019).

To assess the spatial similarities between the shapes of each source, we first ran a overlay analysis using all datasets. The aim was to obtain the distribution of agreement or consensus regarding the spatial extent of Alfama. Then, we computed three different metrics on all pairwise combinations of datasets: intersection over union (IoU), Hausdorff distance and Frechet distance.

While IoU provides a measure of the overlap extent between shapes, the Frechet and Hausdorff distances are common curve similarity metrics. Applied to our case, the former is the smallest of the maximum pairwise distances between vertices of different shapes and the latter is the maximum distance found between a vertex of a shape and the nearest point in the other shape (Bogoya et al., 2019; Lyu et al., 2021). Lastly, we used the shape of the two old parishes as the estimated "ground-truth" of Alfama to determine, for each dataset, values of F-scores and distances towards its centroid.

4 Results and Discussion

The A-DBSCAN output shapes for all datasets are displayed in Fig. 2 while Fig. 3 shows the spatial agreement between sources (including OSM) and the former parishes of São Miguel and Santo Estêvão. The resulting shapes for Alfama varied significantly between datasets, reflecting underlying differences from the sources.

The keyword extraction resulted in extents that indicate the nature of each platform. For instance, Alfama yielded by AirBnB data clearly extends to areas outside the neighbourhood. As expected, Alfama was named at places in other areas, as the website offers touristic short-term rentals. In fact, we opted not to select instances where the keyword was in the description field as the output spread throughout the whole city, due to Alfama being one of the main touristic areas of Lisbon and thus frequently mentioned on the listings. An additional example

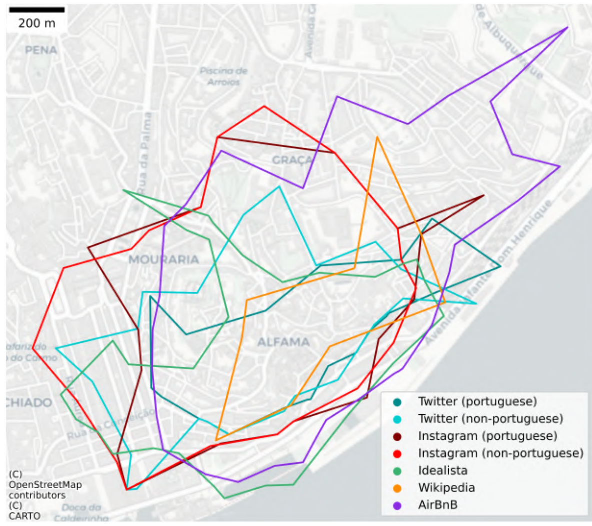

Figure 2. Shapes of the A-DBSCAN results for each dataset with the exception of OSM

is represented by the non-portuguese datasets of Twitter and Instagram. While AirBnB extends to residential areas towards the northeast, the social media outputs stretch to the west, covering parts of the downtown where numerous landmarks and tourist attractions are found.

By observing the outputs' spatial agreement (Fig. 3), we see that a portion of the former parishes overlaps with the highest agreement area. We can also observe that no sources yielded shapes that covered the region alongside the shore, a result of lack of data. Although the agreement provides a general representation of the neighbourhood distribution, UGC data does not directly portray people's perceptions on its extent as opposed to traditional survey methods. Therefore, representations are constrained to users' online activity which are oftentimes bounded to POIs. Nonetheless, administrative units are not the best in depicting neighbourhood boundaries as thought by people. Indeed, Alfama's core does not lie on the shore, but on the hills where it was originally built (Cocola-Gant and Gago, 2021).

In Fig. 4, we take a closer look into the output shapes with the addition of the area retrieved from OSM. Both Instagram and Twitter definitions have their non-portuguese datasets covering larger areas, most noticeable towards the west across downtown. Nevertheless, shapes from Twitter are considerably smaller in extent compared to Instagram. As for Idealista, its Alfama definition also spreads towards downtown, differing substantially from AirBnB. Distinctions suggest that users might be more careful when mentioning Alfama in their real-estate listings when properties are not obviously located in the neighbourhood. On the other hand, AirBnB hosts would benefit more for referencing Alfama on their listings' titles.

Corresponding to the same extent as Google Maps', the OpenStreetMap shape for Alfama does not resemble the former parishes' area. As a collaborative mapping plat-

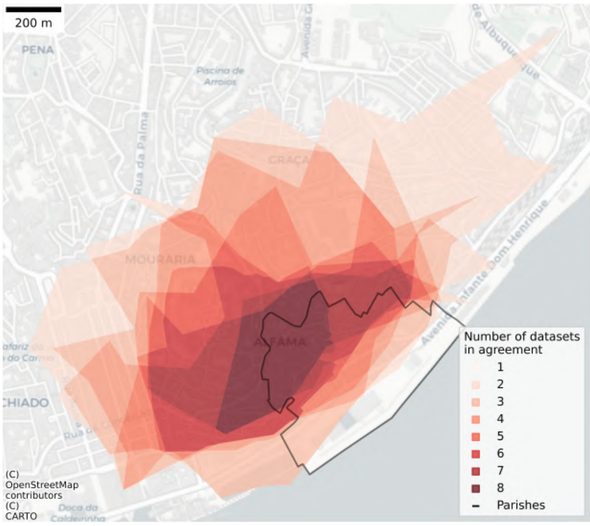

Figure 3. Spatial agreement between datasets and the area corresponding to São Miguel and Santo Estêvão parishes.

Table 1. Results for precision, recall, F-score and distance to centroid (d.c.) of the datasets against the former parishes (São Miguel and Santo Estêvão).

	Recall	Precision	F-score	d.c. (m)
Twitter (PT)	0.27	0.35	0.31	**238**
Twitter (non-PT)	0.17	0.36	0.23	418
Insta (PT)	0.16	0.51	0.24	424
Insta (non-PT)	0.13	0.51	**0.21**	**494**
Idealista	0.31	0.71	**0.43**	339
Wikipedia	0.31	0.32	0.32	**238**
AirBnB	0.16	0.69	0.27	363
OSM	0.21	0.27	0.23	413

form, it is interesting to observe that not only there was only one shape for Alfama in OSM, but also that its area diverges substantially in comparison with the other datasets, although it is covered or intersected by them. The different geographies of Alfama and their spatial agreement reinforces that neighbourhoods are vague entities and individual opinions (in our case, user activity) shows their variable nature, which is in accordance with similar research approaches (Brindley et al., 2018; Twaroch et al., 2019). Nonetheless, diverse levels of spatial cohesion between datasets can provide us a better view of where a certain region is considered to be "more" Alfama than others, alike the work of Montello et al. (2003) for downtown Santa Barbara, California.

We show the measurements of F-score, recall, precision and distance to centroid on Tab. 1 whereas results corresponding to values of IoU, Frechet distances and Hausdorff distances are shown in Tab. 2, 3 and 4 respectively (Appendix). Except for comparisons between same sources but different languages (Instagram and Twitter), AirBnB and portuguese Instagram scored the highest values for IoU. Compared to Twitter, Instagram might have a higher tourism and leisure-oriented online activity, explaining its higher spatial agreement with the AirBnB geography. The similarity of portuguese and non-portuguese geographies from Instagram with AirBnB suggests that visitors are both foreigners and Portuguese or portuguese speakers.

For both Frechet and Hausdorff distances, AirBnB differed the most against the others. Once again, the geography of Alfama for the short-term accommodation platform is defined by a larger extent. In the context of tourism, tolerable walking distances and easier transport accessibility to Alfama might be enough for users to cite the neighbourhood on their listing title. On the other hand, the real-estate website Idealista had the overall shortest Hausdorff distances to other datasets. In other words, the Idealista geography of Alfama is closer to the spatial "average" of all shapes. We did not expect this output from this source, as real-estate listings often exaggeratedly stretch their definition of a neighbourhood for advertising purposes (Twaroch et al., 2019).

Lastly, the lowest and highest values of F-score and distances to the centroid of the former parishes are highlighted in Tab. 1. Compared to the "ground-truth", Idealista yielded the highest F-score (0.43) while non-portuguese Instagram the lowest (0.21). The latter geography also scored the furthest centroid distance towards the old parishes, indicating that foreign users' activity on the social media network collectively refers to the neighbourhood with the lowest accuracy against the extent used by some researchers (Cocola-Gant and Gago, 2021; Madeira et al., 2021). Wikipedia, together with the portuguese Twitter, scored the lowest distances towards the parishes' centroid (238m). The online encyclopedia and the user activity of locals and residents on Twitter seem to be decent proxies of the formal definition of Alfama used in this paper. As mentioned previously, a significant portion of the former parishes' area does not represent the neighbourhood's core.

The OSM definition of Alfama is not the best at representing the neighbourhood, yet it is the definition used by Google Maps, one of the most used geographic information retrieval platforms. While we acknowledge that making discrete spatial judgments regarding neighbourhoods do compromise their vague and fuzzy nature, sharp boundaries are easier to implement in systems that are restrained to visual or computational crisp representations. In addition, web-based sources once again assert themselves as being adequate for inferring people's perceptions on neighbourhoods. We believe that adding more sources of data would certainly improve the consensus on Alfama's spatial definition.

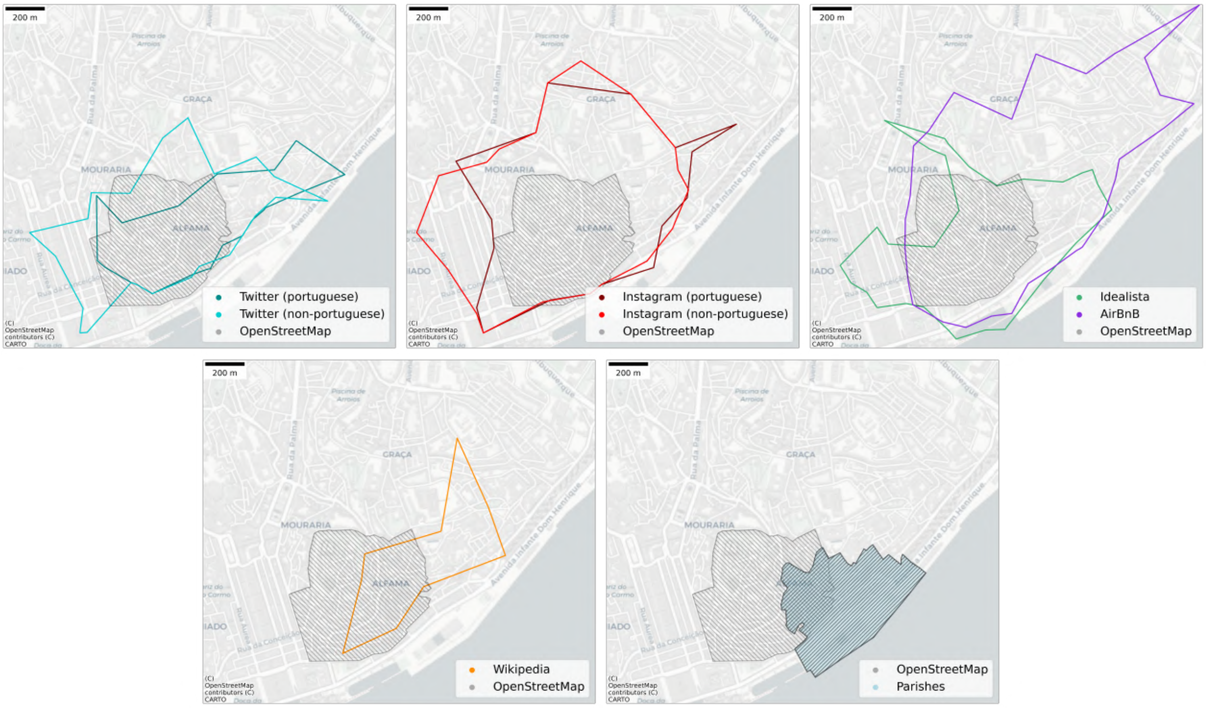

Figure 4. Shapes of the A-DBSCAN results, OpenStreetMap geometry for Alfama and the extent of São Miguel and Santo Estêvão parishes.

5 Conclusions

The results show that the geographies of a historic neighbourhood without official administrative borders will vary depending on the web-based source of information. We demonstrated how different social media networks and websites have different profiles regarding users' activities, intentions and purposes, hence the varying spatial definitions of Alfama. Nonetheless, the combination of several UGC sources and the use of A-DBSCAN provided a straightforward methodological approach aimed at better extracting the discrete extent of the neighbourhood. Obtaining crisp boundaries might not be the best solution for studies on urban perception and spatial cognition, but are useful for defining units for urban planning practices, such as tourism management and resource allocation. A true or ultimate spatial definition is theoretically and practically impossible, however, research that uses neighborhoods as fundamental spatial units of analysis would benefit from alternative definitions that better portray people's perceptions, activities and opinions. Therefore, future work is mainly aimed at refining the approach presented here by retrieving data from more sources, as well as applying and comparing other clustering algorithms and spatial representations. Lastly, asking human participants and examining spatiotemporal patterns (e.g., tourist seasons, COVID-19) would ultimately enhance the delination of the region.

Code and Data Availability

The data collection procedure is outlined at the GitHub repository found here. Lisbon's municipality boundary was retrieved from the city's open geodata portal and the former distribution of parishes (before 2012) was collected from the website of the Portuguese national land management bureau. As for OSM, we exported the data using the extent of the city and filtered the Alfama neighbourhood shape. AirBnB listings data was downloaded from the Inside Airbnb project website.

We carried out the data pre-processing and analysis through Python language on Jupyter Notebook as well as QGIS software environments. All data and codes supporting this paper are available on a FigShare repository

Acknowledgement

The authors appreciatively acknowledge the funding from the Portuguese national funding agency for science, research and technology (Fundação para a Ciência e a Tecnologia - FCT) through the CityMe project (EXPL/GES-URB/1429/2021; https://cityme.novaims.unl.pt/).

CRediT authorship contribution statement

Vicente Tang: Conceptualization, Methodology, Software, Validation, Writing - Original Draft, Visualization.

Jaskaran Puri: Software, Data Curation, Methodology, Writing - Review and Editing. **Marco Painho**: Writing - Review and Editing, Supervision.

Competing interests

The authors declare that there are no competing interests.

References

Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P., Joho, H., and Sanderson, M.: Web-based delineation of imprecise regions, Computers, Environment and Urban Systems, 30, 436–459, https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2005.08.001, 2006.

Arribas-Bel, D., Garcia-López, M.-, and Viladecans-Marsal, E.: Building(s and) cities: Delineating urban areas with a machine learning algorithm, Journal of Urban Economics, 125, 103 217, https://doi.org/https://doi.org/10.1016/j.jue.2019.103217, delineation of Urban Areas, 2021.

Bogoya, J. M., Vargas, A., and Schütze, O.: The Averaged Hausdorff Distances in Multi-Objective Optimization: A Review, Mathematics, 7, https://www.mdpi.com/2227-7390/7/10/894, 2019.

Brindley, P., Goulding, J., and Wilson, M. L.: Generating vague neighbourhoods through data mining of passive web data, International Journal of Geographical Information Science, 32, 498–523, https://doi.org/10.1080/13658816.2017.1400549, 2018.

Catney, G., Frost, D., and Vaughn, L.: Residents' perspectives on defining neighbourhood: mental mapping as a tool for participatory neighbourhood research, Qualitative Research, 19, 735–752, https://doi.org/10.1177/1468794118803841, 2019.

Chen, M., Arribas-Bel, D., and Singleton, A.: Understanding the dynamics of urban areas of interest through volunteered geographic information, Journal of Geographical Systems, 21, 89–109, https://doi.org/10.1007/s10109-018-0284-3, 2019.

Cocola-Gant, A. and Gago, A.: Airbnb, buy-to-let investment and tourism-driven displacement: A case study in Lisbon, Environment and Planning A: Economy and Space, 53, 1671–1688, https://doi.org/10.1177/0308518X19869012, 2021.

Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N.: The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City, 2012.

Cunha, E. and Martins, B.: Using one-class classifiers and multiple kernel learning for defining imprecise geographic regions, International Journal of Geographical Information Science, 28, 2220–2241, https://doi.org/10.1080/13658816.2014.916040, 2014.

Daly, P., Dias, A. L., and Patuleia, M.: The Impacts of Tourism on Cultural Identity on Lisbon Historic Neighbourhoods, Journal of Ethnic and Cultural Studies, 8, 1, https://doi.org/10.29333/ejecs/516, 2021.

Davies, C., Holt, I., Green, J., Harding, J., and Diamond, L.: User Needs and Implications for Modelling Vague Named Places, Spatial Cognition & Computation, 9, 174–194, https://doi.org/10.1080/13875860903121830, 2009.

de Oliveira, T. H. M. and Painho, M.: Open Geospatial Data Contribution Towards Sentiment Analysis Within the Human Dimension of Smart Cities, pp. 75–95, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-58232-6_5, 2021.

Galster, G.: On the Nature of Neighbourhood, Urban Studies, 38, 2111–2124, https://doi.org/10.1080/00420980120087072, 2001.

Gao, S., Janowicz, K., and Couclelis, H.: Extracting urban functional regions from points of interest and human activities on location-based social networks, Transactions in GIS, 21, 446–467, https://doi.org/https://doi.org/10.1111/tgis.12289, 2017a.

Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J.-A., McKenzie, G., Ju, Y., Gong, L., Adams, B., and Yan, B.: A data-synthesis-driven method for detecting and extracting vague cognitive regions, International Journal of Geographical Information Science, 31, 1245–1271, https://doi.org/10.1080/13658816.2016.1273357, 2017b.

Goodchild, M. F.: Citizens as sensors: the world of volunteered geography, GeoJournal, 69, 211–221, https://doi.org/10.1007/s10708-007-9111-y, 2007.

Hollenstein, L. and Purves, R.: Exploring place through user-generated content: Using Flickr tags to describe city cores, Journal of Spatial Information Science, 1, 21–48, 2010.

Hu, S., Xu, Y., Wu, L., Wu, X., Wang, R., Zhang, Z., Lu, R., and Mao, W.: A framework to detect and understand thematic places of a city using geospatial data, Cities, 109, 103 012, https://doi.org/https://doi.org/10.1016/j.cities.2020.103012, 2021.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S.: Extracting and understanding urban areas of interest using geotagged photos, Computers, Environment and Urban Systems, 54, 240–254, https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2015.09.001, 2015.

Jones, C. B., Purves, R. S., Clough, P. D., and Joho, H.: Modelling vague places with knowledge from the Web, International Journal of Geographical Information Science, 22, 1045–1065, https://doi.org/10.1080/13658810701850547, 2008.

Lyu, C., Wu, X., Liu, Y., and Liu, Z.: A Partial-Fréchet-Distance-Based Framework for Bus Route Identification, IEEE Transactions on Intelligent Transportation Systems, pp. 1–6, https://doi.org/10.1109/TITS.2021.3069630, 2021.

Madeira, A., Palrão, T., Mendes, A. S., and López-Morales, E.: Perceptions about Tourism and Tourists in Historic Neighborhoods: The Case of Alfama, Sustainability, 13, https://doi.org/10.3390/su13158357, 2021.

Mai, G., Janowicz, K., Hu, Y., Gao, S., Zhu, R., Yan, B., McKenzie, G., Uppal, A., and Regalia, B.: Collections of Points of Interest: How to Name Them and Why it Matters, in: Spatial big data and machine learning in GIScience, Workshop at GIScience 2018, edited by Raubal, M., Wang, S., Guo, M., Jonietz, D., and Kiefer, P., pp. 29–33, Leibniz International Proceedings in Informatics, Melbourne, Australia, 2018.

Montello, D. R., Goodchild, M. F., Gottsegen, J., and Fohl, P.: Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries, Spatial Cognition & Computation, 3, 185–204, https://doi.org/10.1080/13875868.2003.9683761, 2003.

Montello, D. R., Friedman, A., and Phillips, D. W.: Vague cognitive regions in geography and geographic information science, International Journal of Geographical Information Science, 28, 1802–1820, https://doi.org/10.1080/13658816.2014.900178, 2014.

Papadakis, E., Resch, B., and Blaschke, T.: Composition of place: towards a compositional view of functional space, Cartography and Geographic Information Science, 47, 28–45, 10.1080/15230406.2019.1598894, 2020.

Rêgo, C. S. and Almeida, J.: A framework to analyse conflicts between residents and tourists: The case of a historic neighbourhood in Lisbon, Portugal, Land Use Policy, 114, 105 938, https://doi.org/https://doi.org/10.1016/j.landusepol.2021.105938, 2022.

Schockaert, S., De Cock, M., and Kerre, E. E.: Automatic Acquisition of Fuzzy Footprints, in: On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, edited by Meersman, R., Tari, Z., and Herrero, P., pp. 1077–1086, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

Sequera, J. and Nofre, J.: Touristification, transnational gentrification and urban change in Lisbon: The neighbourhood of Alfama, Urban Studies, 57, 3169–3189, https://doi.org/10.1177/0042098019883734, 2020.

Shang, S., Guo, D., Liu, J., Zheng, K., and Wen, J.-R.: Finding regions of interest using location based social media, Neurocomputing, 173, 118–123, https://doi.org/https://doi.org/10.1016/j.neucom.2015.06.086, 2016.

Tang, V., Acedo, A., and Painho, M.: Sense of place and the city: the case of non-native residents in Lisbon, Journal of Spatial Information Science, pp. 125–155, https://doi.org/10.5311/josis.2021.23.165, 2021.

Tulumello, S. and Allegretti, G.: Articulating urban change in Southern Europe: Gentrification, touristification and financialisation in Mouraria, Lisbon, European Urban and Regional Studies, 28, 111–132, https://doi.org/10.1177/0969776420963381, 2021.

Twaroch, F. A., Brindley, P., Clough, P. D., Jones, C. B., Pasley, R. C., and Mansbridge, S.: Investigating behavioural and computational approaches for defining imprecise regions, Spatial Cognition & Computation, 19, 146–171, https://doi.org/10.1080/13875868.2018.1531871, 2019.

Appendix

Table 2. Intersection over Union (IoU) between pairwise combinations of UGC data sources.

	Twitter (non-PT)	**Insta (PT)**	**Insta (non-PT)**	**Idealista**	**Wikipedia**	**AirBnB**	**OSM**
Twitter (PT)	0.48	0.33	0.28	0.37	0.48	0.29	0.45
Twitter (non-PT)		0.54	0.54	0.47	0.32	0.37	0.49
Insta (PT)			0.79	0.40	0.26	0.58	0.39
Insta (non-PT)				0.41	0.21	0.50	0.33
Idealista					0.29	0.39	0.42
Wikipedia						0.24	0.27
AirBnB							0.30

Table 3. Frechet distances (m) between pairwise combinations of UGC data sources.

	Twitter (non-PT)	**Insta (PT)**	**Insta (non-PT)**	**Idealista**	**Wikipedia**	**AirBnB**	**OSM**
Twitter (PT)	1041	595	611	1022	709	1318	626
Twitter (non-PT)		652	1360	375	1148	1667	707
Insta (PT)			792	664	1071	1667	668
Insta (non-PT)				1313	725	952	730
Idealista					1155	1604	719
Wikipedia						794	942
AirBnB							1465

Table 4. Hausdorff distances (m) between pairwise combinations of UGC data sources.

	Twitter (non-PT)	**Insta (PT)**	**Insta (non-PT)**	**Idealista**	**Wikipedia**	**AirBnB**	**OSM**
Twitter (PT)	354	595	611	395	355	847	626
Twitter (non-PT)		367	307	375	642	1040	517
Insta (PT)			360	436	605	678	668
Insta (non-PT)				453	724	937	593
Idealista					598	1000	392
Wikipedia						794	519
AirBnB							1260