A social media-based framework for tourist behaviour analysis and characterization in urban environments

Francisco Porras-Bernardez *, Georg Gartner

Research Group Cartography, TU Wien, Vienna, Austria. francisco.porras.bernardez@tuwien.ac.at, georg.gartner@tuwien.ac.at

* Corresponding author

Abstract:

Tourism is a very important and fast growing industry worldwide that has generated 25% of all global net new jobs during the last 5 years. New tools can be valuable for relaunching the sector and provide alternative analysis and segmentation capabilities to organizations involved. We present an analysis and visualization framework for tourist behaviour study and segmentation based on tested methods and technologies, combined and extended in an innovative way. Our framework uses Flickr data as input and classifies users according to country of origin. Then, urban distribution patterns are obtained in two different spatial levels by using [Network] Kernel Density Estimation in 1D and 2D spaces, as well as spatial clustering with HDBSCAN. Basic Natural Language Processing is applied to extract and visualize semantics generated in the social media platform and a visualization of typologies of Points of Interest by nationality is proposed for the development of tourism dashboards. We have applied our framework to three European cities of different size to test the segmentation capabilities of the approach. Results suggest a good potential for tourism management in urban environments.

Keywords: Geo-Social Media, Tourism, Flickr, KDE, HDBSCAN, Cultural Heritage

1. Introduction

Tourism is a major source of income for national economies worldwide and has experienced a continuous growth and diversification during the last decades. It is one of the largest economic sectors in the planet and, despite its huge environmental impact, it is opening new development opportunities for more and more regions (Dwyer et al., 2020). As of 2019, tourism induced impact in the world represents 7.5€ billion, a 10.3% global GDP and a 10.4% of total employment. In Europe, tourism contributes with 1.7€ Billion, 9.1% of GDP and 37.1 million jobs (*Global Economic Impact & Trends 2020*, 2020).

The current situation generated by Covid-19 has impacted most of the economic sectors in the world. Tourism is one of the most affected areas due to a fundamental dependence in mobility and safety. Nevertheless, innovative tools for tourism study and promotion are still necessary for a future recovery of a sector that has generated 25% of all global net new jobs during the last 5 years (*Global Economic Impact & Trends 2020*, 2020).

According to (J. Li et al., 2018) tourism research using big data is relatively novel; during the last 13 years a maximum of 30 annual publications can be found. However, the scientific production is speeding up. These authors classify research in three primary sources: user-generated content (UGC), devices (GPS, Bluetooth, WiFi, etc.) and operational data (web searches, web visits, credit cards, etc.). In tourism research, UGC accounted for almost half of the literature. UGC geolocated is also known as Volunteered Geographic Information (VGI).

One of the most interesting sources of VGI is Social Media (Kitchin, 2013). Some of social media platforms include the geographic component by allowing the users to indicate geolocation on their posts and/or themselves (user's current location). In the literature, these Social Media are referred to as Geo-Social Media (GSM). GSM data has been used in studies focused on different fields such as events detection (Walther & Kaisser, 2013; Xu et al., 2020), mobility analysis (Hawelka et al., 2014; Pan et al., 2013; Rashidi, 2017), user behaviour (Yang et al., 2017) or urban planning (Ciuccarelli et al., 2014).

There have been identified multiple limitations of GSM as a source of reliable data. The main drawbacks are related to quality and quantity of the data. Different coverage of urban vs. rural areas (Hecht & Stephens, 2014), popularity of the platform (Williams et al., 2016) or data amount offered by the medium itself (Leetaru et al., 2013), are among the most common issues mentioned in the literature. This influences the representativeness of the samples and often has been argued that there are biases regarding the population depicted in terms of socioeconomic characteristics (Graham et al., 2014). Nevertheless, it has also been presented as a useful tool for the study of urban environments for different purposes such as disaster management (Granell & Ostermann, 2016), travel behaviour (Rashidi et al., 2017) or tourism promotion (Hvass & Munar, 2012), among many others.

The work presented in this paper was developed within the framework of the project Eureca: EUropean Region Enrichment in City Archives and collections (Verstockt et al., 2019) of the Technical University of Vienna (Research Group Cartography), the University of Ghent (IDLab,

CartoGIS), and several state and city archives from Vienna and Ghent. Eureca worked on revealing traces (i.e. influences or links) of European regions that have shaped the settlements in which we live today. These tracks are conformed by historical, cultural, economic, political and architectural aspects, which can be used as anchor elements to disclose cultural heritage items related to specific European territories. The project developed tools to explore such traces when visiting an urban settlement. Eureca aimed to encourage visitor's discovery of cultural heritage linked to specific regions or nationalities. The enriched metadata generated during the project is usable to develop further research, foster the exploitation of archive collections to a wider public and lure new groups of cultural heritage consumers. Thus, we developed tools potentially useful for tourism promotion.

As one of the information sources to identify European traces, we chose GSM because is a modern, dynamic and ever-growing source of information. A huge amount of spatially-related data with a large geographic and social coverage is available. In contrast, traditional survey techniques are costly and can be very limited in terms of spatial coverage. On the other hand, we have already mentioned some limitations that make GSM not a flawless mirror of society. Therefore, any analysis based on this source of information has to be taken with an adequate dose of caution.

The main innovation of this work lies on the novel combination and extension of existing methods to generate a general framework for the analysis and visualization of behaviour of tourists in urban environments, in terms of movement patterns.

In the following sections, the whole method will be thoroughly described and a use case scenario will be presented and analyzed.

2. Method

Our method involved the combination of several different techniques in subsequent steps logically connected (see Figure 1). Firstly, we prepared the data to build a geo-analysis-ready big dataset. Then, the users from the GSM were classified by country of residence with the support of a classification algorithm. We aimed to study spatial behaviour patterns of different nationalities at a general and a more specific level, and within an urban environment. For the "*general*" level, we focused on larger areas and modelled *footprints* by applying Kernel Density Estimation (KDE). For the more "*specific*" level, we modelled Areas of Interest (AOIs) and Streets of Interest (SOIs) by using spatial clustering and Network Kernel Density Estimation (NKDE) respectively. In a subsequent phase, map algebra helped us to estimate overall differences in user presence in different city areas.

A semantics' analysis using basic Natural Language Processing (NLP) was carried out to extract Topics of Interest (TOIs): i.e. the most mentioned terms in their posts. In a final phase, well-known Points of Interest (POIs) of the city were compared with the footprints so as to build tourist profiles according to nationality.

We used *Spyder*, an open-source scientific Python development environment, to build most of our framework.
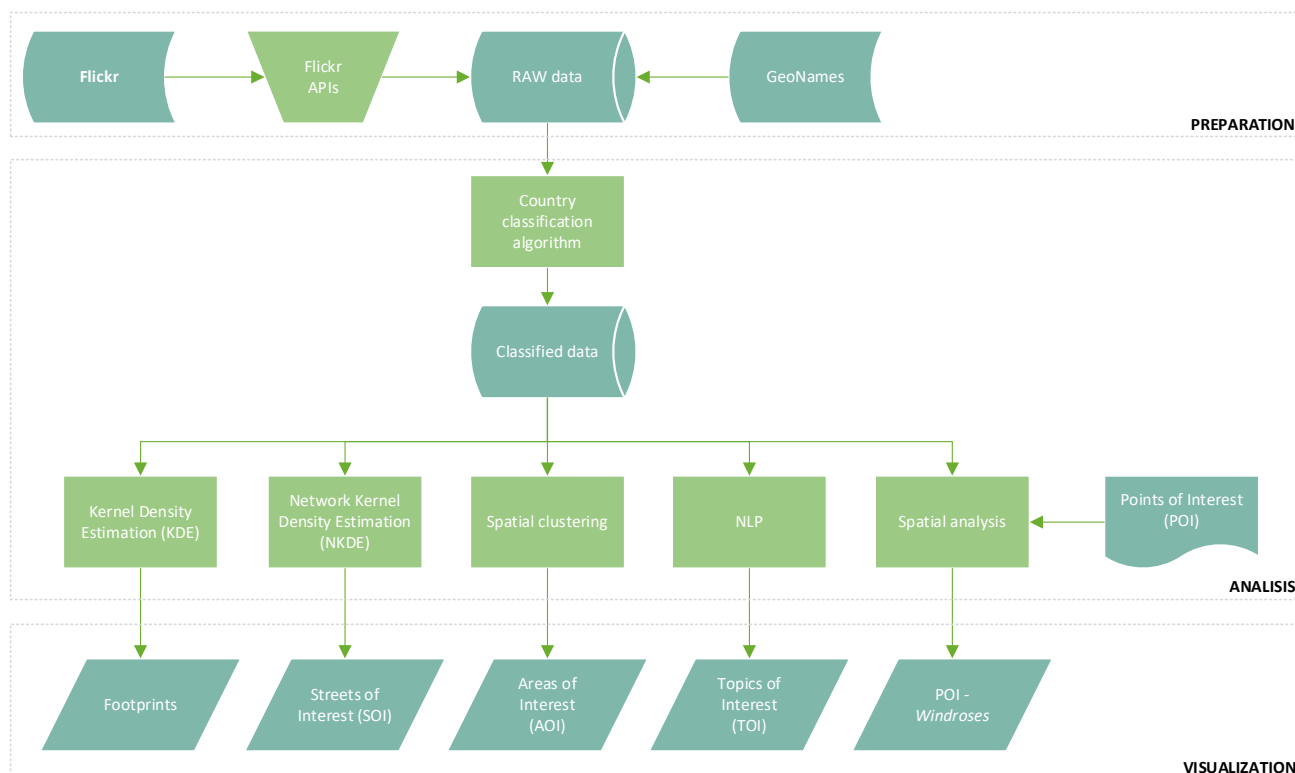

Figure 1. Framework diagram.

2.1 Data preparation

In an initial phase, we collected data from Flickr. The basic setup consisted in several scripts for data collection and processing, and one PostgreSQL spatial database.

Despite there are several relatively big public datasets available, we decided to build our own dataset in order to overcome problems such as limited spatial and temporal coverage, reduced number of attributes or limited accessibility. Otherwise, we would have had to limit the area of study, variety of nationalities, temporal representativeness and reach of the modelling. Moreover, we aimed at contributing our own large dataset open for further research.

Two of the public Application Programming Interfaces (APIs) offered by Flickr were used to retrieve metadata of pictures. The attributes of the picture included ID, uploading date, title, tags and geolocation. Additionally, data about the owner/user was collected, i.e. ID, name and self-reported location. This location was in a heterogeneous format, so if only a city name was provided, the GeoNames gazetteer (www.geonames.org) was used to determine to which country the city belongs.

An area of 66 M km2 centred in continental Europe was defined for data collection. The API has limitations for picture retrieval. Searches can be done spatially by using a bounding box, but a maximum of 4K unique photos (metadata) are delivered by each search. Any extra pictures are duplicates. To overcome this limitation, the 66 M km2 area was divided in square cells small enough to guarantee less than 4K pictures in the output JSON file. Then, our script used these cells as a sliding window to call the API enough times to cover all Europe. Moreover, a search has to indicate a time period, so each cell was retrieved hundreds of times to cover the whole time of interest. Our script dynamically adjust the temporal coverage of the search between 3 and 7 days to keep the number of results below the 4K limit.

The final temporal range of our dataset covered 2004-2018 and the amount of pictures (metadata) collected reached 68 million. Our aim is to release the dataset to the public soon, after adapting it to the sharing policies of the platform.

2.2 Country of residence determination

For any use in tourism studies, it is fundamental that the users are differentiated at least by country of residence. In our dataset, only around 32% of the users indicated their place of residence in their profiles. Hence, in order to be able to make use of most of the Flickr posts for spatial behaviour analysis, a user's country of residence classification algorithm was needed. For the sake of simplicity, the terms "nationality", "country of origin" and "home location" are used in the rest of this paper as synonyms for "country of residence" of the GSM user. The focus of this research is on analysing tourists depending on the country they are coming from in a relatively short-term perspective (as long as the data in the social media platform). Our aim was not to achieve this classification by true nationality or place considered "home" by the user, given that this is virtually impossible based on the kind of raw data stored in the SM platform.

A new algorithm for country of origin classification was developed based on literature (L. Li & Goodchild, 2012; Paldino et al., 2015) and implemented in Python. Basically, all the pictures uploaded by the user in all the countries of analysis were considered. Then, from those countries in which the user uploaded pictures for a period greater than 6 months, the country with highest number of uploads was selected as home location.

After applying this algorithm, around 75% of the Flickr contributors finally had information about their country of residence. In order to evaluate the quality of the classification, this algorithm was benchmarked using as ground truth data the European users that originally included home location information in their profiles. The algorithm reached a precision of 87% and a recall of 76%. Based on previous works (Da Rugna et al., 2012), it can be considered that the performance of the algorithm is acceptable. Once the users were classified by country of origin, it was possible to carry out spatial analysis to determine distribution patterns of tourists from different countries.

2.3 General patterns: *footprints*

The geolocation of the pictures was stored as points in our database. In order to visualize them in a meaningful way, we generated continuous raster surfaces by using *Kernel Density Estimation (KDE)* (Grothe & Schaab, 2009) with ArcGIS (ArcMap). Each raster is a heatmap depicting areas of variable concentration of pictures and representing a footprint of the visitors in the city. Given that the pictures' owners were classified by country of origin, specific footprints were modelled for different nationalities. Thus, it was possible to visualize an estimation of the urban areas attracting more visitors from
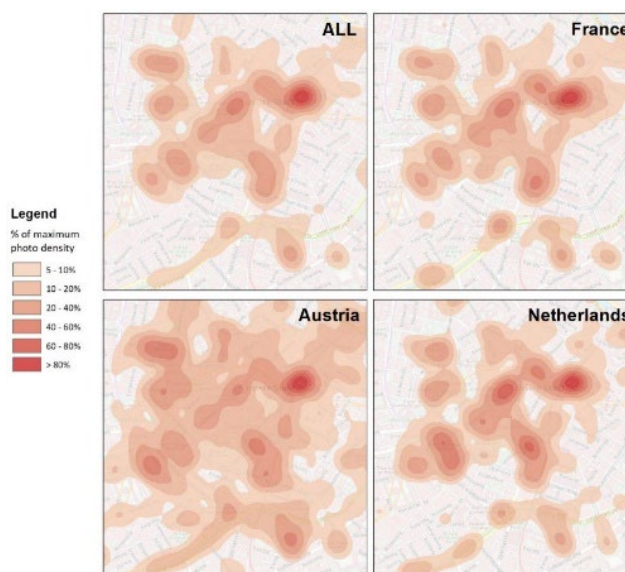

Figure 2. Vienna KDE footprints of visitors from several countries of origin.

individual countries, as depicted in Figure 2.

Applying KDE required the selection of a grid *cell size* and a search radius or *bandwidth*. The cell size had to be enough to resolve the limits of the region and we set it according to the city size and bandwidth calculated. The adequate bandwidth is commonly determined in an experimental way (Yu et al., 2015), however in our work we estimated an optimal bandwidth (Zhou et al., 2018) for each nationality dataset in each city. This takes into consideration the sample size and the standard distance as a measure of dispersion of each point distribution.

An overall footprint depicted the general distribution of tourists in the city without differentiation of country of origin. By using Map Algebra (Tomlin, 1994), the national footprints were compared with the overall footprint with the purpose of mapping genuinely attractive or unattractive areas for specific nationalities. Here attractiveness represents a high or low density of pictures taken by the visitors, relative to the overall number of pictures taken by the group under consideration. Additionally, the national footprints can be compared against each other to identify areas more or less visited for one country compared to the other (see Figure 4).

In the next step, we reduced the spatial level to focus our analysis on smaller areas, more related to urban structures such as neighbourhoods, streets, squares, etc.

2.4 Specific patterns

At a lower level of analysis, spatial patterns more "specific" were modelled. At this level, the objective were smaller hot spots with higher concentration of pictures, hence representing areas that attract people's attention. In a twofold approach, the focus was on Areas of Interest (AOIs) and Streets of Interest (SOIs).

AOIs

They represented smaller areas in the city surface, considered as a 2D space. In order to model these areas, spatial clustering was performed on the pictures with the algorithm HDBSCAN (Campello et al., 2013). Then, a convex hull was generated for each obtained cluster so that, at the end, a collection of polygonal surfaces were depicting the urban AOIs. For these tasks, a script making use of the ArcGIS ArcPy library was implemented.

HDBSCAN is an improved version of the density-based DBSCAN algorithm (Ester et al., 1996), transforming it in a hierarchical algorithm in order to extract a group of density-based clusters. A single parameter is required to be set for HDBSCAN, this is the minimum number of points per cluster (*minPts*). In order to establish and adequate value for *minPts*, a simple strategy was followed. First, an ideal number of AOI (n) was manually determined for each city, according to data from the official regional tourism organizations. Second, clustering tests were performed for each nationality dataset with a range of *minPts* values representing 0.1-2% of the total points for each nation in each city. Third, the *minPts* value was chosen according to A) generating the smallest variation between the maximum and minimum number of clusters obtained for each nationality dataset and B) being the closest to n.

SOIs

The second focus was on the street network as a 1D space. In an urban environment users move mainly through the streets; therefore, we decided to model also footprints of visitors along this network as a complement to the other modelling. Points distributed throughout a network are typically analysed with spatial methods that assume Euclidean distance. However, Euclidean distances and their equivalent short-path distances are significantly different (Okabe & Sugihara, 2012). In previous works, Network Kernel Density Estimation (NKDE) has also been used (Delso et al., 2018; Okabe et al., 2006) for the estimation of the density of points on a network. The ArcGIS toolbox SANET (Okabe et al., 2018) was used in our analysis.

Both AOIs and SOIs could be of help in tourism or urban planning by showing areas and streets (Figure 3) with a recorded higher attraction of tourists, thus revealing potential targets for tourism promotion measures. Moreover, these measures could be tailored to specific nationalities. After these steps targeting tourist spatial behaviour, the focus of interest changed to the semantics generated by tourists associated to different spaces.

2.5 Semantic analysis

NLP techniques are often used for the study of the words more prevalent in semi-structured textual content.

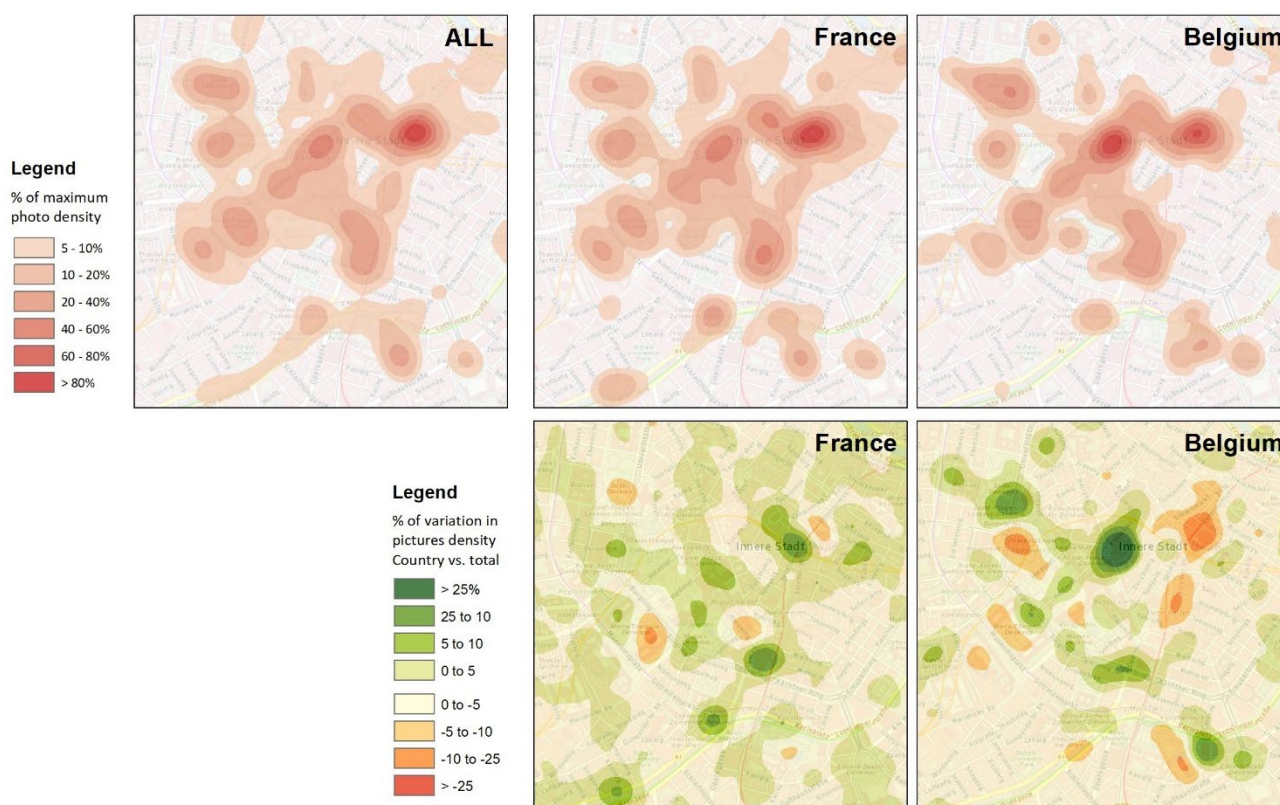Figure 3. Vienna NKDE street footprints of visitors.

Figure 4. Vienna KDE footprints after map algebra.

A simplistic approach using NLP was followed to analyse and visualize the most frequent terms mentioned in pictures' metadata. These can be named as Topics of Interest (TOIs). A script in R making use of the *tm* and *igraph* R packages was used to clean and count frequency of appearance of the terms. The first part of the processing involves the cleaning of the picture tags from special characters, stopwords, terms related with commercial brands, years and specific terms referring to photography such as *black&white*, etc. Then, a corpus is created with all the clean tags and a term-document matrix is generated, considering each Flickr picture as a document. The most popular words are selected: those with a frequency of appearance greater than a threshold. This threshold was between the 92 and 99 percentile depending on the number of pictures per nationality. Multiple tests were performed for each country of origin in order to select the most "equilibrated" graph for that country, i.e number and size of nodes, complexity and readability of the graph.

Then graphs were created with nodes representing the terms and edges reflecting the relations between them. The node size is proportional to the total frequency of appearance and the thickness of the edge is relative to the frequency of co-appearance of both nodes in a picture. This approach allows visualizing the terms more used and how often they are used together in pictures for a specific area and by a particular tourist group.

In the previous steps, spatial and semantic analyses were performed in order to generate insight into the spaces and semantics contained in the dataset and related with the physical world. Then, we worked on the opposite direction

and tried to relate the physical world with the dataset by considering urban items that include cultural heritage elements. These items were taken as POIs and we aimed to identify how different categories of items attract different visitor groups by creating tourist profiles for the main nationalities in the dataset.

2.6 Tourists profiling

A collection of the main POIs in three cities was gathered. The popularity of the POIs was determined according to tourism-related datasets offered by official open data portals (www.geopunt.be, www.data.gv.at) in combination with the popular travel and tourism website of *Tripadvisor* (www.tripadvisor.com). These POIs were classified in eight categories according to assumed main interests for a tourist: *shopping/transport, live events, sights, museums, religious, historical buildings, green areas and gastronomy*. This classification was not intended to be a comprehensive and well-tested taxonomy, but just a base to test this tourist profiling approach. The following step relied on the national KDE footprints. The KDE surfaces were divided in intervals of 10% of density of pictures. For the purposes of this paper, those POIs laying within areas of more than 50% of maximum density were considered relevant enough for the interest of the visitors. This threshold was chosen as a middle value representing areas in which pictures were taken with at least half of the maximum intensity. Then, *wind-rose* graphs were built for each tourist nationality (see Figure 7). Each POI category is represented in one "arm" of the graph and the size of the arm is proportional to the percentage of total POIs visited for this category.

3. Case study

In order to test our framework, we selected three European cities: Vienna, Ghent and Brussels. Vienna and Ghent represent two different sizes of settlements. We added Brussels as being the capital of the European Union, hosting a wide variety of nationalities as well as receiving visitors from all over the world. The footprints modelled showed differences according to the country of origin of the visitors. It was possible identifying diffuse areas of special concentration of pictures for some nationalities.

When comparing the overall tourist **footprint** with each of the national models, it was possible to identify areas of genuine attraction or rejection for specific nationalities in comparison with the average tourist.

When dealing with the lower level patterns, the spatial clustering generated more compact AOIs. As reflected by Figure 5, clear areas of special attraction for some tourist segments became obvious. When the focus was on the street network, the linear footprints generated by the NKDE discovered genuine patterns of street segments more used by visitors from specific nations.
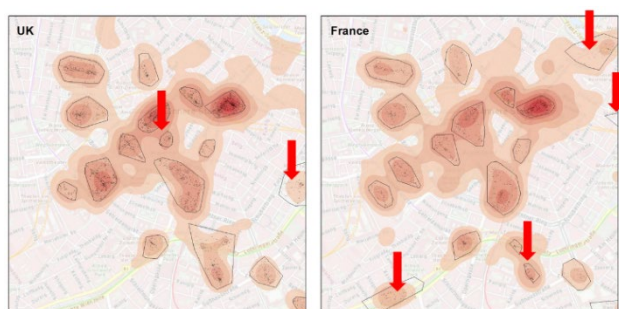

Figure 5. Vienna KDE footprints and AOIs: Comparison of visitors from France and UK.

The semantic analysis of the pictures metadata allowed us to obtain the TOIs as graph visualizations for the different tourist groups. Terms easily linkable to specific nationalities and languages popped-out in the graphs. Moreover, cultural heritage-related terms were more often present for some countries of origin. Figure 6, shows a topic graph for visitors from France containing terms such as *artnouveau, baroque, église, ottowagner*, etc.

Finally, the POI profiles were built for the main countries of origin of the tourists visiting each city (see Figure 7). These combined visualizations provided a quick insight on the typology of venues the tourists are interested on when visiting each city.

4. Conclusions

This paper presented a methodology that builds on top of existing techniques in an innovative way and proposes an integrated framework for tourist behaviour analysis and visualization. Different analysis methods have been used to identify spatial patterns that represent movement behaviour in different scales and types of space. Moreover, we have identified semantic patterns that reflect the tourist interests within different areas of the city. This framework

produces a tangible output comprised of a set of multiple representational items such as *footprints, areas, streets, topics* and *points* of **interest**. Our work offers a planning tool for tourism segmentation integrating all the processes from data collection to visualization.

Several limitations have been mentioned like the aspects in representativeness of Social Media, the dependence on the accuracy of the home determination algorithm or the subjectivity of the classification of tourist activities in the different venues. Nevertheless, we consider this is an initial valid step in establishing a well-defined tourism analysis methodology based on low cost, highly accessible alternative data sources. Moreover, we are proposing a new tool to foster storytelling in Geovisualization.
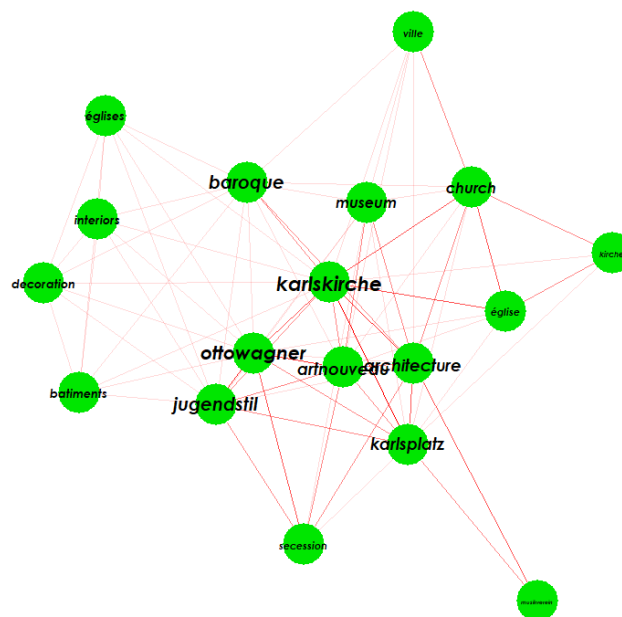

Figure 6. Topic graph for TOIs for France in the *Karlsplatz* area (Vienna).

Figure 7. Typologies of POIs visited by some tourist nationalities in Vienna

5. Open challenges and future directions

Further research will be needed to improve the home classification algorithm as well as a new benchmarking with a more complete dataset. Other social media might be included and compared with Flickr. The coverage of the dataset should be enlarged to include contributions from all over the world. This could very likely improve the

quality of the classification and the variety of countries of origin modelled.

Moreover, additional KDE tests with more bandwidths could be beneficial. Further improvements could include a thorough urban POI classification, as well as a more complex semantic analysis including sentiments. Finally, the visualizations could be integrated in a web-based dashboard with a subsequent user testing process.

6. Acknowledgements

7. References

Campello, R. J., Moulavi, D., & Sander, J. (2013). *Density-based clustering based on hierarchical density estimates*. 160–172.

Ciuccarelli, P., Lupi, G., & Simeone, L. (2014). *Visualizing the data city: Social media as a source of knowledge for urban planning and management*. Springer. http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-8031

Da Rugna, J., Chareyron, G., & Branchet, B. (2012). Tourist behavior analysis through geotagged photographies: A method to identify the country of origin. *2012 IEEE 13th International Symposium on Computational Intelligence and Informatics (CINTI)*, 347–351. https://doi.org/10/ghjkz6

Delso, J., Martín, B., & Ortega, E. (2018). A new procedure using network analysis and kernel density estimations to evaluate the effect of urban configurations on pedestrian mobility. The case study of Vitoria –Gasteiz. *Journal of Transport Geography*, *67*, 61–72. https://doi.org/10/gdj29r

Dwyer, L., Forsyth, P., & Dwyer, W. (2020). *Tourism Economics and Policy* (2nd ed.). Channel View Publications, 2020.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise. 96*(34), 226–231.

Global Economic Impact & Trends 2020. (2020). World Travel & Tourism Council.

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the World are You? Geolocation and Language Identification in Twitter. *The Professional Geographer*, *66*(4), 568–578. https://doi.org/10/gc6zmp

Granell, C., & Ostermann, F. O. (2016). Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management. *Computers, Environment and Urban Systems*, *59*, 231–243. https://doi.org/10/ghjpvt

Grothe, C., & Schaab, J. (2009). Automated Footprint Generation from Geotags with Kernel Density Estimation and Support Vector Machines. *Spatial Cognition & Computation*, *9*(3), 195–211. https://doi.org/10/b776jj

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*(3), 260–271. https://doi.org/10/f3sqg3

Hecht, B., & Stephens, M. (2014). A Tale of Cities: Urban Biases in Volunteered Geographic Information. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 197–205.

Hvass, K. A., & Munar, A. M. (2012). The takeoff of social media in tourism. *Journal of Vacation Marketing*, *18*(2), 93–103. https://doi.org/10/gjkv3x

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267. https://doi.org/10/gc7rdb

Leetaru, K., Wang, S., Guofeng, C., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat, The geography of Twitter.pdf. *First Monday*, *18*(5). https://doi.org/201305191201

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Li, L., & Goodchild, M. F. (2012). Constructing places from spatial footprints. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information - GEOCROWD '12*, 15. https://doi.org/10/ggfmpw

Okabe, A., Okunuki, K., & SANET Team. (2018). *SANET. A Spatial Analysis along Networks (Ver.4.1)*.

Okabe, A., Okunuki, K., & Shiode, S. (2006). SANET: A Toolbox for Spatial Analysis on a Network. *Geographical Analysis*, *38*(1), 57–66. https://doi.org/10/b7tbjg

Okabe, A., & Sugihara, K. (2012). *Spatial analysis along networks: Statistical and computational methods*. Wiley.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, *4*(1), 5. https://doi.org/10/gdz5tr

Pan, B., Zheng, Y., Wilkie, D., & Shahabi, C. (2013). Crowd sensing of traffic anomalies based on human mobility and social media. *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 344–353. https://doi.org/10/gfwtsx

Rashidi, T. H. (2017). *Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges*. 15.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*, 197–211. https://doi.org/10/f9q8bk

Tomlin, C. D. (1994). Map algebra: One perspective. *Landscape and Urban Planning*, *30*(1–2), 3–12. https://doi.org/10/dm2qm2

Verstockt, S., Milleville, K., Ali, D., Porras-Bernardez, F., Gartner, G., & de Weghe, N. V. (2019). *EURECA - EUropean Region Enrichment in City Archives and collections*. 9.

Walther, M., & Kaisser, M. (2013). Geo-spatial Event Detection in the Twitter Stream. In P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, & E. Yilmaz (Eds.), *Advances in Information Retrieval* (pp. 356–367). Springer. https://doi.org/10/ghj99x

Xu, Z., Liu, Y., Yen, N. Y., Mei, L., Luo, X., Wei, X., & Hu, C. (2020). Crowdsourcing Based Description of Urban Emergency Events Using Social Media Big Data. *IEEE Transactions on Cloud Computing*, *8*(2), 387–397. https://doi.org/10/ghj99q

Yang, L., Wu, L., Liu, Y., & Kang, C. (2017). Quantifying Tourist Behavior Patterns by Travel Motifs and Geo-Tagged Photos from Flickr. *ISPRS International Journal of Geo-Information*, *6*(11), 345. https://doi.org/10/ghj992

Yu, W., Ai, T., & Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography*, *45*, 32–47. https://doi.org/10/f7ggvh

Zhou, H., Li, Y., & Jiang, T. (2018). Sea Clutter Distribution Modeling: A Kernel Density Estimation Approach. *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, 1–6. https://doi.org/10/gk9j5n