Location-Based Social Network Data for Exploring Spatial and Functional Urban Tourists and Residents Consumption Patterns

Aurélie Cerdan Schwitzguébel
CETT-UB School of Tourism, Hospitality & Gastronomy
Barcelona, Spain

Oriol Romero Bartomeus
Clarivate Analytics
Barcelona, Spain

Submitted: 5th December 2018; Resubmitted: 13rd December 2018; Accepted: 21st December 2018. e-ISSN: 2014-4458

Abstract

Urban tourist destinations' increasing popularity has been a catalyst for discussion about the tourist activity geographical circumscription. In this context, Big Data and more specifically location-based social networks (LBSN), appear as a valuable source of information to approach tourist and residents spatial interactions from a renewed perspective.
This paper focuses on approaching similarities and differences between tourists and residents' geographical and functional use of urban economic units. A user classificatory algorithm has been developed and applied on YELP's Dataset for that purpose. A residents and tourists integration ratio has then been calculated and applied by types of businesses categories and their associated spatial distribution of the of 11 metropolitan areas provided in the sample: Champaign (Illinois, US), Charlotte (North Carolina, US), Cleveland (Ohio, US), Edinburgh (Scotland, UK), Las Vegas (Nevada, US), Madison (Wisconsin, US), Montreal (Quebec, CA), Pittsburgh (Pennsylvania, US), Phoenix (Arizona, US), Stuttgart (DE) and Toronto (Ontario, CA). Business category results show strong similarities in tourists and residents functional coincidence in the use of urban spaces and leisure offer, while there is a clear geographical concentration of activity for both user types in all analysed case studies.

Resumen

La creciente popularidad de los destinos urbanos ha actuado como catalizador del debate sobre la delimitación geográfica de la actividad turística. En este contexto, el Big Data, y más específicamente las redes sociales que integran ubicación (LBSN), aparecen como una valiosa fuente de información para aproximarse a la interacción espacial entre turistas y residentes, desde una perspectiva renovada.
Este artículo se centra en la aproximación a las similitudes y diferencias entre el uso geográfico y funcional de las unidades económicas urbanas, por parte de turistas y residentes. Para ello, se ha desarrollado y aplicado un algoritmo de clasificación de usuarios a un conjunto de datos de YELP. Se ha calculado también un ratio de integración entre turistas y residentes urbanos, posteriormente aplicado a los negocios georreferenciados y sus categorías funcionales, en las 11 áreas metropolitanas incluidas en la muestra: Champaign (Illinois, EEUU), Charlotte (Carolina del Norte, EEUU), Cleveland (Ohio, EEUU), Edimburgo (Escocia, GB), Las Vegas (Nevada, EEUU), Madison (Wisconsin, EEUU), Montreal (Quebec, CA), Pittsburg (Pennsylvania, EEUU), Phoenix (Arizona, EEU), Stuttgart (DE) and Toronto (Ontario, CA).
Las categorías funcionales que agrupan los negocios muestran claras similitudes en cuanto a la coincidencia espacial entre turistas y residentes. Además, hay una clara concentra-

Key Words:
Urban Tourism, Big Data, Yelp, Spatial Analysis, Consumption Patterns.

ción geográfica de la actividad para ambos grupos de usuario en todos los casos estudiados.

Palabras clave:
Turismo Urbano, Big Data, Yelp, Análisis Espacial, Patrones de Consumo.

Introduction. Problem Statement

The importance of tourist activity in urban destinations has been increasing for several years and has become a global trend, raising new challenges for planners and destinations managers. It has been argued that the tourist activity, if not well managed, can contribute to the loss of multifunctionality of urban spaces that receive a higher pressure (García-Hernández, de la Calle-Vaquero, & Yubero, 2017). On the basis of this statement, multiple theories have appeared that seek to model the urban tourism phenomenon from a spatial and functional perspective. Yet the complex nature of cities as dynamic systems (Fernández Güell & López, 2016) complicates obtaining precise information about the geographical distribution of urban tourists. Big data is then seen as a new opportunity which, through the emergence of new techniques and methodologies that allow its obtaining, processing and use, has allowed a widening of the resources focus and new approximations to the urban tourism phenomenon.

The present research presents the initial exploration of YELP's open-access dataset and seeks to set the base to further develop a methodology oriented to the identification of the integration of tourists and residents in an urban destination. To achieve this, exploratory research from a dual functional and spatial perspective has been carried out. The research is then addressed from the consumptive nature of the tourist activity, under the assumption that any type of business and resource can have tourist potential.

Objectives

The main objective of the present research is to analyse the integration of urban tourists and residents consumption patterns in the different urban environments provided by YELP's dataset. This integration is more specifically understood as the proportional amount of interaction, or user-generated content, that urban tourists and residents create and link with the same economic units.

To achieve that, two specific objectives have been formulated: a) to analyze the integration level of tourists and residents by functional businesses categories, by comparing the registered interaction (namely reviews posted) with the different type of businesses in each analysed city, and b) to analyze the spatial and temporal evolution of tourists and residents integration level in different urban environment, by visualizing the annual evolution of the registered interaction for all businesses types and each city.

Theoretical Framework

Spatial clustering of tourism as an economic activity

The traditional inclusion of commercial activities and infrastructure in the consideration of "tourism resources" highlights how strongly entwined the consumerist nature of tourism is with the development, planning and management of a destination. Aside from historically important or primary resources such as accommodation and historic heritage, these secondary resources and infrastructure support are key components of the tourism system (Vera Rebollo, López Palomeque, Marchena Gómez, & Anton Clavé, 2011), thus supporting the main attractions (Burtenshaw, Bateman, & Ashworth, 1991).

The involvement of complementary services overcomes the imaginary geographical and functional division between the tourist functions and the rest of the destination. This aspect of the destination's evolution has aroused and still does the interest of scholars and experts. In that way, the urban space is often classified depending on the level of the predominance of the tourist function and phenomenon on others (Burtenshaw et al., 1991; Getz, 1993b; Hayllar & Griffin, 2005; Hayllar, Griffin, & Edwards, 2008; Jansen-Verbeke, 1998; Pearce, 1998).

Functional clustering conceptualisations have evolved since Ashworth and Tunbridge's (cited in Pearce, 1998) definition of "entertainment districts", which highlighted the tendency to agglomerate of catering facilities and nightlife venues, as well as to tourist-oriented facilities such as hotels and tourist attractions. The later formulation of Burtenshaw et al., (1991) Central Tourist District (CTD) and Recreational Business District was likewise supported by the idea of the embedding of the tourist activities in the city, and that these tourists will limit their spatial consumption of the urban spaces to specific attractions and services' concentrations and linking corridors. Getz, (1993) Tourist Business District (TBD) definition recovered this idea and nuanced the inner-city tourism system conceptualization composed by "activity place" and "leisure setting" (Jansen-Verbeke, 1986), as being delimitated complementary spaces to the general leisure and recreational urban space. Therefore, TBD assimilates the notion of spatially concentrated tourist-oriented attractions and supply facilities while it reflects on the multi-motivational nature of urban tourists. TBD will then spatially coincide or overlap with the Central Business District (CBD), as well as include a diversified range of components, namely conferences and exhibitions venues, heritage and historical attractions, waterfronts, museums and art galleries, shopping offer, catering facilities, entertainment, theatres and concert halls, sports venues and facilities, viewpoints and connections. (Judd, 1999) subsequently develops the idea of the "tourist bubble" based on the analysis of several case studies, which leads to the identification of spaces segregated from the rest of the city, in which facilities and amenities oriented to the tourist trade tend to agglomerate. Here again, those facilities include catering businesses, sports venues, nightlife offer, but also souvenir shops. The description of tourist districts, on the other hand, assimilates the notions of social and cultural heterogeneity and economic multi-functionalism (Jansen-Verbeke & Ashworth, 1990; Pearce, 2001) in the identification of six non-exclusive districts types: historic districts, ethnic districts, sacred spaces, redevelopment zones, entertainment destinations and functional tourist districts (inspired in Getz's TBD definition). This classification underlines the variety of urban tourist destinations, moves away from the traditional and dominant sectoral view, but also takes into account the importance of the management strategies undertaken by each city. Along with this line, the concept of tourist precincts emerge as a distinct geographic area differentiated of its surroundings by its variety and type of supply facilities, land uses, and the presence of a singular physical feature that contributes to visitors' interest (Hayllar & Griffin, 2005; Hayllar et al., 2008). There is a visible evolution in this conceptualisation of precincts, in the fact that they are conceived not only as tourist areas of interest but as of interest and enjoyment of residents too.

Tourism, and hence consumption, have progressively influenced the spatial and economic structures of post-modern cities as tourist destinations (Ashworth & Page, 2011; Hayllar et al., 2008a; Judd, 1999; Page, 1995; Urry, as cited in Page & Hall, 2003). Urban environments have then partially evolved into consumption and productions spaces (Pearce, 2001), where tourism and leisure offer is supplied by specialised types of equipment (Rowe and Stevenson, as cited in Page & Hall, 2003). It is widely accepted that urban tourists are multi-motivational (Ashworth & Page, 2011; Burtenshaw et al., 1991; Hayllar et al., 2008; Rogerson & Rogerson, 2016; Vera Rebollo et al., 2011) which, in combination with the attractiveness of the specialised offer available in urban destinations, implies a non-exclusive use of resources, infrastructures and services whose original function was non-touristic in nature (Ashworth & Page, 2011; Burtenshaw et al., 1991; Shaw and Williams , as cited in Page & Hall, 2003). Motivated by the similarities with tourists' interests, resident's consumption patterns concurrently adapt to this evolution of the commercial fabric and network. This spatial overlapping of functions and types of demand causes a blurring the physical and functional boundaries of the complex supply network (Britton, as cited in Judd, 1999).

With that in mind, the idea of a tourist-resident interaction space emerges (Edwards, Griffin, & Hayllar, 2008; Hayllar et al., 2008). Ashworth and Page, (2011) return to this idea when noting that "the 'tourist city' could only be conceived alongside and overlapping with, other 'cities'". The areas of the tourist city will then be integrated by attractions and supply facilities which attract tourists' attention, without necessary displacement of other functions of the urban spaces. In fact, the loss of the aforementioned multi-functionality will signify the creation of urban resorts similar to Judd's (1999) "tourist bubbles".

In the manner that proximity, accessibility, land rent, comparative shopping, existing infrastructure, investment and regeneration policies, and the concentration of other components are key drivers of tourism economic units clustering tendency (Ashworth & Page, 2011; Pearce, 1998), tourists as consumers demonstrate several differences in their consumption patterns (such as the shortness of their stay) that will directly affect the concentration of tourist flows (Lew & McKercher, 2006; Rogerson & Rogerson, 2016; Shoval & Raveh, 2004; B. Zhou, Tang, Zhang, & Wang, 2014; Zhou, Xu, & Kimmons, 2015). Paradoxically to the fact that the tourist activity needs of the existence of a supply infrastructure to develop, the "new urban tourists" tends to avoid typically categorised "tourist places" and are more and more attracted by the "ordinary life of a city" (Füller & Michel, 2014). That endless quest for authenticity, as discussed by MacCannell, (1976/2017) is not a new phenomenon, nevertheless the growing popularity of urban destinations and its strong relation with the consumption of the urban culture, lifestyle or identity (Judd, 1999; Kannisto, 2018; Page & Hall, 2003), represents a new challenge for urban planners and managers.

Big Data in urban tourism research

Urban destinations' growing affluence of tourists and the similarity between tourist and residents consumption patterns has been a traditional obstacle in the analysis of the real impact of tourism in the urban environment (Kádár, 2013). Traditional studies forms have proven to be limited in obtaining socio-spatial and temporal tourist and residents behaviours, thus complicating the planning and management of urban destinations (Florido Trujillo, Garzón García, & Ramírez López, 2018; Salas-Olmedo, Moya-Gómez, García-Palomares, & Gutiérrez, 2018). In this context, the high cost of customary research tools as surveys or interviews, the increasing data availability and the growing focus on data management and analyse methods have motivated the exploitation of big data potential in the tourism research field (Maeda, Yoshida, Toriumi, & Ohashi, 2018). Considering that classic socio-spatial tourist data collection has often been narrowed to specific aspects of the destinations at the expense of others, for example omitting less popular attractions or excursionists-related data, the use of big data has gained relevance as a complementary or alternative source of information. For instance, Leung, Vu, & Rong, (2017) were unable to find official statistics about non-first-tier attractions and had complemented their study using user-generated content.

The analytical applications of big data, which is characterised by the massive volume of diverse information contained in datasets, by the high-speed in its generation from varied sources, and by the sophistication of the analytical and management technologies and systems that it requires (Katal, Wazid, & Goudar, cited in Batista e Silva et al., 2018; Gandomi and Haider, cited in Marine-Roig & Anton Clavé, 2015; McAfee & Brynjolfsson, cited in Önder, 2017) are diverse and versatile. Accordingly, there has been eclosion of tourist literature linked with computational science and the use of big data sources, namely user-generated sources, that has grown constantly since 2007 (J. Li, Xu, Tang, Wang, & Li, 2018).

This emphasis placed in user-generated content (UGC) as a big data source in tourism research is strongly associated with the interest aroused by the digital footprint left by both tourist and residents in their interaction with the destination's components (Önder, 2017; Salas-Olmedo et al., 2018; Scherrer, Tomko, Ranacher, & Weibel, 2018). This interaction includes online textual data, images and videos, all actively provided by users and of relative ease of access and, most importantly, georeferenced information (Kuo, Chan, Fan, & Zipf, 2018; J. Li et al., 2018). And more so, Social Networks Sites (SNS) have become the hallmark source of web-based user-generated content, where individuals and businesses create profiles and share information and knowledge (Sapountzi & Psannis, 2016), even though the characteristics of big data itself (and even more those of online social network data) as dynamic, and massive often unstructured amount of data, can potentially hamper analytical research (Sapountzi & Psannis, 2016; Zhou et al., 2015). Other obstacles of using social networking sites, microblogs, community media sites, location-based social networks and messaging platforms as sources of information are the implicit bias caused by the differences in behaviours depending on the type of users of each platform; the duplication in user-counting when using more than one SNS information source at the same time; the difficulties related to processes of extraction of information itself such as the lack of structure and its noisy nature; the lack of data availability; as well as content trust issues (J. Li et al., 2018; Maeda et al., 2018; Pranata & Susilo, 2016; Salas-Olmedo et al., 2018; Sapountzi & Psannis, 2016; Zhou et al., 2015).

Even taking into account these difficulties, the use of geographic information systems (GIS) to approach socio-spatial user behaviour through location-based social networks (LBSN) platforms is increasingly relevant, in particular by means of check-in data (Twitter) or reviews linked to geo-tagged venues (Foursquare, Tripadvisor or Yelp) (Stock, 2018; Zhou et al., 2015). This is explained by the range of possibilities that the availability of dynamic user-centred georeferenced information and context metadata offers in the analysis of the socio-spatial behaviour and patterns in the urban environment, such as mapping segmented variables, and identifying points of interest (POI) or areas of interest (AOI) (García-Palomares, Gutiérrez, & Mínguez, 2015; Leung et al., 2017; Maeda et al., 2018; Önder, 2017; Shao, Zhang, & Li, 2017), and its application in the fields of urban destinations marketing and management. POI and AOI detection, as the identification of places that generate interest and affluence of visitors or hotspots, have been one of the central research aims of computer science application to tourism research. Identifying not only the geographical concentration of visitors, but also contextual information such as interests, opinions, temporary-distribution of visits, or nationalities, is marking significant contributions to the identification and parameterisation of urban tourism areas or districts, and allows an alternative approach to the geographical distribution of the urban space multifunctional function.

Related work

POIs and AOIs clustering mapping is a common application of LBSN georeferenced data extraction and facilitates the visual display of complex, massive amount of information. Zhou et al. (2015) demonstrated the applicability of geospatial analysis based on cloud computing in their urban tourist hotspot identification based on Flikr geotagged photos, as well as the feasibility in the obtaining of popular associated tags and keywords of such clusters. Similarly, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is also used by Shao et al. (2017) when mapping natural, recreational and cultural urban tourist districts in the Huangshan City in China. In this latter case, the data that allows the detection of tourist communities is fetched from Sina-Weibo, and based on the assumption that interactions recorded at tourism-related POIs are more likely to belong to tourists.

Salas-Olmedo et al. (2018) conducted an extraction of Madrid tourist AOIs that emphasises the possibilities of combining multiple SNS data in the analysis of tourists' digital footprint. Their study, based on Panoramio, Twitter and Foursquare data, concludes that central urban spaces tend to be more multifunctional than peripheral ones, as tourists' use of the city is temporally and spatially limited.

Brandt, Bendler, & Neumann (2017), in their research about the relation between tweet semantics location and topic engagement with tourist destinations, further argue that data density is not exclusive of tourist areas but also of residents' AOIs. Maeda et al. (2018) have applied a modified DBSCAN algorithm in combination with Term-Frequency Inverse Document Frequency (TF-IDF) method to infer differences in domestic (Japanese) and international tourist POIs preferences. Their findings shed light on the possibilities of applying SNS data sources to segment consumption and behavioural patterns in the urban environment.

García-Palomares et al. (2015) mapped visually attractive POIs in several European cities and analysed the spatial distribution differences between tourist and residents georeferenced photographs activity. They concluded that the geodata distribution radius is more concentrated in the case of tourists', but that there is a spatial overlap in the case of the city's most representative sites. Similarly, Mukhina, Rakitin, & Visheratin (2017) developed a methodology to segment SNS dataset users as tourists and residents to further identify popular hotspots for both groups in Saint Petersburg. Their research, based on geotagged Instagram's entries, detected differences in the activity temporal distribution, more specifically that tourist displayed a higher level of activity during summer weekdays (coinciding with the tourist high-season), but also significant overlapping during the rest of the year. In line with the foregoing, Li, Zhou, & Wang (2018) centred their analysis in the spatial interaction between tourists and residents. The focus on spatial integration and segregation between the two groups of users is especially relevant, and their findings of a variable level of integration not only between cities but also across the same city. In this study again, tourist hotspots appear to be more concentrated, while residents consume scattered POIs.

Methodology

Yelp Datasety

This research, exploratory in nature, is based on the dataset provided on the occasion of Yelp's 2018 11th Round annual challenge under open-access conditions for academic purposes (retrieved from https://www.yelp.com/dataset/challenge). The dataset originally contained a total of 174.567 businesses records, 1,3 million users and 5,2 million reviews, has been previously filtered by Yelp's recommendation system, and covers 11 metropolitan areas. That recommendation filter implies that only businesses with 3 reviews older than 14 days are included, even if the total reviews published for each user (*review_count* variable) is also available. This allows a comparison of the total reviews posted by each user with the reviews included in the metropolitan areas analysed. The data is structured up into six main groups: Business, Check-ins, Reviews, Tips, Users and Photos. This research is based on Business, Reviews and User information only, which can be interrelated through the *business id*, *review id* and user id variables.

Data cleaning and normalisation

The data was provided in JSON format and has been inserted into a MongoDB database to interrelate the selected objects (Businesses, Check-ins, Photos, Reviews and Users).

Metropolitan areas contiguity

The dataset contained semi-structured data from unknown metropolitan areas and multiple city variables. To ensure correct plotting of the main cities with their immediate surrounding administrative settlements, a grid-based approach has been undertaken. Standard zooms 10 (20.480 metres wide tile side), 15 (640 metres wide tile side) and 18 (80 metres wide tile side) have been applied to divide the metropolitan areas in smaller square shaped tiles. A list of tiles associated with at least 1 business has been built. Metropolitan area boundaries have then been initially defined by the plotting of tiles at zoom 10 associated to at least one business, and the subsequent grouping of the contiguous tiles. The segregation of tiles whose 8 contiguous tiles have been found empty, has allowed the initial identification of secondary settlements. A second plotting phase from a tile at zoom 15 has allowed refinement of these boundaries. Specifically, the tile containing the higher number of businesses has been identified and used as a reference in plotting tiles by contiguity.

That process allowed the removal of businesses incorrectly attributed to specific cities and to ensure that georeferenced coordinates matched the metropolitan areas to which they were associated according to the city field. As a result, 168.506 businesses distributed in 11 metropolitan areas were identified, most of them located in the United States (US) and Canada (CA): Champaign (Illinois, US), Charlotte (North Carolina, US), Cleveland (Ohio, US), Edinburgh (Scotland, UK), Las Vegas (Nevada, US), Madison (Wisconsin, US), Montreal (Quebec, CA), Pittsburgh (Pennsylvania, US), Phoenix (Arizona, US), Stuttgart (DE) and Toronto (Ontario, CA).

Tag feature business classification

Because of the focus on tourist and residents interaction, all purely tourist-oriented businesses tags were removed from the sample, including all tourist accommodation-tagged, souvenir shops and other highly-specialised resort-type not emplaced in urban environments (some of which are ski or golf resorts). Several tags were considered too general to allow a correct classification and consequently ignored. The remaining 1087 categories tags are non-exclusive and

Table 1 Grouped Business Functional Categories

Monuments, landmark & heritage	Shops and stores
Museums, art galleries	Offices and diverse work premises
Cinemas, concert venues & theatres	Sports venues and related services
Nightclubs, bars & nightlife offer	Public mobility infrastructures & services
Cafes, bars, restaurants & catering activities	Private transport services

Source: Prepared by the author.
Adapted from Burtenshaw, Bateman, & Ashworth, (1991).

allowed a spatial overlap of multiple functions to better reflex the nature of urban space, through the application of a modified version of the ground theory sorting proposed by Burtenshaw et al. (1991), as listed in Table 1.

Data Processing

Residents identification

Related studies which incorporate distinctions between tourists and residents in their methodology were based on users' profiles arguing that this would avoid erroneous segmentations (Li, Zhou, & Wang, 2018). However, this information was not included in the dataset and is the reason an algorithm base on the work of Mukhina et al. (2017) and García-Palomares et al. (2015) was developed. The following assumptions are made and based on previous research and case studies:

- Urban tourists' stay at the destination is shorter than 2 weeks. The average length of stay for the analysed cities is shorter than one week in all cases (12 days if considering stay in the whole region) according to official statistical data of the responsible authorities (Arizona Office of Tourism, 2016; Commonwealth of Pennsylvania, 2015; IHK Region Stuttgart, 2017; Institut de la Statistique Québec, 2014; Las Vegas Convention and Visitor Authority, 2017; Visit North Carolina, 2017; Visit Scotland, 2016). Previous studies have considered lapses between 1 week (Salas-Olmedo et al., 2018) and 30 days (Mukhina et al., 2017).

- Urban tourists won't return to the destination in a 6 months period.

- Users that post reviews to a specific business have to have visited it.

As some users might have posted very few comments in their home city, assuming that they are tourists would significantly affect the research findings. For this reason, intervals between quartiles have been calculated for both user total reviews (*review_count*) and available reviews per user (*review_length*) and cross-referenced to estimate their statistical dispersion and discard less reliable users. The applied algorithm consists of the following several consecutive steps:

1. Users whose total posted reviews (*review_count variable*) equals 2 or less have been discarded, in considering the provided information not being sufficient for a reliable classification.

2. Users whom only 1 review is available (*review_length variable*) have been classified as tourists in the metropolitan areas where that unique available review is geo-referenced.

3. Users with more than 1 available reviews and more than 2 total reviews have been classified according to the temporary window that limits their review activity:

   3.1. Users whose reviews associated to a metropolitan area can be grouped in periods wider than 15 consecutive days have been marked as residents of that metropolitan area, and as tourists in the remaining clusters.
   3.2. Users whose reviews associated to a metropolitan area can be grouped in 2 or more groups of 15 consecutive days, whose time lapse in between of the groups is smaller than 6 months (180 days) have been marked as residents of that metropolitan area, and as tourists in the remaining clusters.

4. Users with less than 25% of available reviews of the total *review_count* but whose reviews are related to business associated with the same metropolitan area have been classified as tourists in that metropolitan area.

5. Users whose 40% total *review_ length* or more are related with business associated with the same metropolitan area have been classified as residents of that cluster.

6. Users with more than 25% of available reviews of the total *review_count*, but whose reviews are related with business associated with the same metropolitan area, have been classified as residents of that cluster.

7. The remaining users who don't fit any of the rules fixed have been dismissed.

As a result, 436.536 users (a 33% of the initial users) have been dismissed after the classificatory process.

Integration ratio

Li et al. (2018) adapted Sakoda's, (1981) Dissimilarity Index proposal, used to obtain a global Index of residents and tourists integration. Following their methodology, an integration ratio R has been calculated to estimate the level of integration between residents and tourists for each of the analysed businesses of the dataset. R ratio is estimated considering the unique (non-duplicated) tourists (t_i) and local residents (l_i) that have posted reviews in each of the venues. Both values are divided by all unique tourists (T) and residents (L) identified for the metropolitan area where the venue is located. The difference in the proportion of tourists ($\frac{t_i}{T}$) and residents ($\frac{l_i}{L}$) interacting with that venue represents the level of interaction between the two groups:

$$R = \frac{t_i}{T} - \frac{l_i}{L}$$

The closer R is to 1, the more important that venue is in terms of tourist affluence and the less in terms of residents presence. On the contrary, venues whose ratio is negative and closer to -1 are proportionally more frequented by residents. The more extreme (setting -1 and 1 as extremes) the ratio is, the less integrated are both groups. Including total unique users for each of the groups when establishing the proportion adds perspective contributes to data normalization, and ultimately to compare cases where tourist affluence can strongly vary in term of absolute affluence numbers. In other words: the tourist activity greatly varies from one destination to another as not all cities receive a comparable number of visitors. Additionally, tourists may tend to comment on businesses or venues more frequently than residents, even if the formers one might visit the business or venues repeatedly. Basic descriptive statistics analysis has been applied to explore functional categories and associated businesses and venues integration ratio by metropolitan areas.

Mapping Evolution/ Spatial clustering

R ratios have then been spatially aggregated for obtaining of Z values. The ratios by tile (Z values) have been obtained by calculating the average R of all businesses contained in each of the tiles (n). Z has been calculated for each of the tiles at zoom 15 (409.600 m^2) and 18 (6.400 m^2) of each metropolitan areas grid-based extension.

$$Z = \frac{1}{n} \sum_{i=1}^{n} R$$

Results and discussion

As hinted above, metropolitan areas present wide differences in users and businesses size of the samples. For instance, Las Vegas is the metropolitan area with the largest sample of users followed by far by Phoenix and Toronto. In all cases, except for Phoenix, Toronto and Charlotte, more tourists than residents have been identified, symbolising the difference in tourism activity scale. Similarly, Phoenix, Las Vegas and Toronto concentrate most of the businesses of the dataset, as displayed in table 2:

Table 2 Absolute frequency of unique users and total businesses by metropolitan areas

Metropolitan Area	Total unique users	Visitors	Residents	Nº Businesses
Champaign	8.688	5.612	3.076	1.618
Charlotte	31.427	5.612	25.815	12.244
Cleveland	43.621	26.266	17.355	10.036
Edinburgh	7.872	5.729	2.143	3.854
Las Vegas	402.050	265.804	136.246	32.378
Madison	22.553	14.193	8.360	4.100
Montreal	31.755	23.016	8.739	7.558
Phoenix	262.629	114.602	148.027	50.399
Pittsburgh	43.898	26.164	17.734	9.438
Stuttgart	5.616	3.303	2.313	2.428
Toronto	73.927	30.306	43.621	28.719

Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017

Business functional distribution comparison by metropolitan areas

In general, there is a clear preponderance of businesses tagged as “Cafes, bars, restaurants & catering activities” (43,92% on average), followed by “Offices and diverse work premises” (36,51%) and “Shops and stores” (28,35%). Even considering the fact that tags have not followed a proportional distribution, there is a clear specialization in catering-related services: of 1087 total tags, only 118 (close to 11%) were assigned to that category. On the contrary, both “Offices and diverse work premises” and “Shops and stores” concentrate 57,29% and 15,51% of all tags, respectively. Phoenix, Las Vegas, Champaign and Charlotte are exceptions in which the businesses distributions show a slightly higher presence of offices and workplaces. The reason of such specialization remains uncertain, as it can be argued whether restaurants and catering-oriented services businesses tend to optimise their online presence, or whether if there is, in fact, such a strong presence of this type of business in all case studies.

“Nightclubs, bars & nightlife offer” category follows by far with an average of 10,17% of all venues classified within this group, except for the case of Edinburgh (18%) mostly due to the inclusion of “pubs” as nightlife-related tag. Paradoxically, Las Vegas proportion of businesses categorised as nightlife-oriented (8%) ranks below the average (10%) even if its absolute frequency is still higher than in any other case with a total of 2.593 businesses.

The remaining categories group less than 4% of all venues in all metropolitan areas, or even less than 1% in the case of “Museums, art galleries”, “Public mobility infrastructures & services”, “Private transport services” and “Monuments, landmark & heritage”.

However, all cities present a spatial overlap of all functional categories which can be explained by the concentration tendency of georeferenced businesses. Besides, functional categories are non-exclusive and most of the analysed businesses are classified in more than one group at the same time. As illustrated by the example of Toronto’s metropolitan area in figures 1, 2, 3 and 4 (at tile zoom 18), all businesses and venues’ distribution concentration follows the urban layout, outlining the presence of important streets and roads. There is a clear spread of offices and working places (that includes approximately 27% of the total Toronto’s businesses) while both cafes and other catering activities (that in the case of Toronto represent more than a 50% of all businesses) and shops and stores (28%) tend to concentrate in smaller areas. Still, both shops and catering businesses clusters clearly overlap and coincide with the same clusters that show higher density economic units. This is represented by the darker shade of red, in contrast to light orange and yellow shades used to represent low-density in businesses by tile (see figures 1, 2, 3 and 4). Functional categories that group a smaller quantity of businesses is found to replicate this pattern at a smaller scale with, in the case of the lowest frequencies, no clear clustering tendency but even then spatial overlapping with other categories.

All metropolitan areas display significant density of businesses in specific clusters with multifunctional orientation. Business sprawl varies accordingly to the business frequency of the different business categories. Logically, categories with fewer businesses (as monuments and landmarks of important mobility infrastructures) don’t display concentration patterns as some of them appear as isolated, Nevertheless, when comparing their locations with another activity type, overlapping is made visible. These results suggest no general functional difference between urban spaces. It should anyway be considered to further segment the categories applied in this research, as some of them (namely offices and work premises) include a wide amount of diverse type of businesses, what can affect the final results.

This distribution pattern is repeatedly reproduced in the remaining ten metropolitan areas analyzed (Figures 1, 2, 3 and 4).[1]

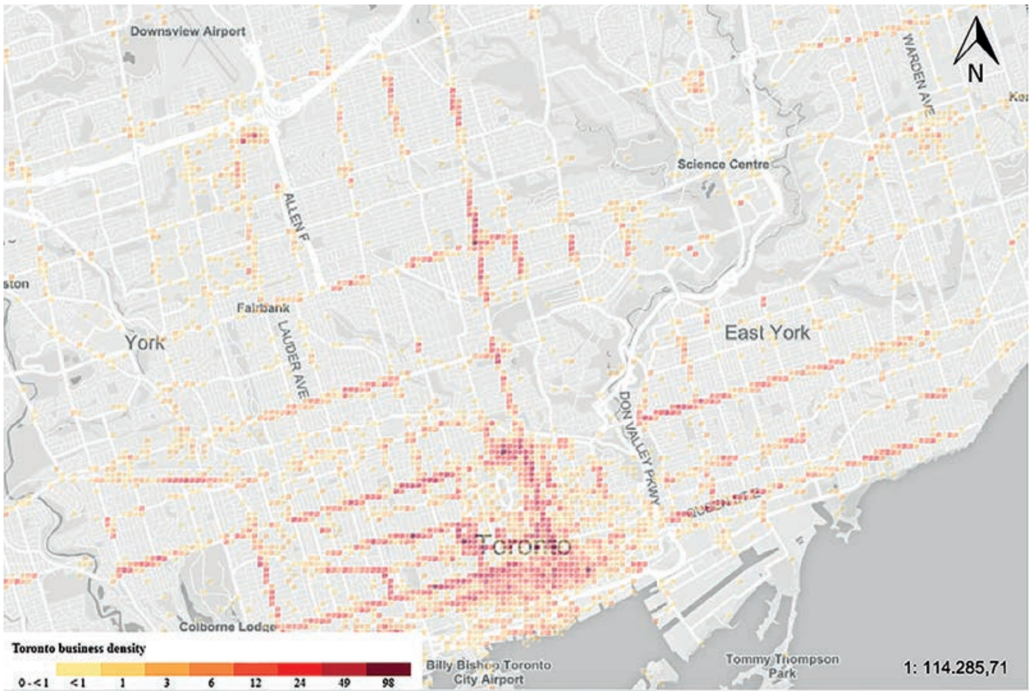
Tourists and residents integration by functional category

At a city level, all integration ratios appear to be negative and extremely close to 0, starting from Edinburgh that appears to have the lowest ratio (-0,00365006513 0042) to Las Vegas and its highest integration level (-0,000034703222347). From Edinburgh to Las Vegas, Champaign ranks 2nd and is followed by Stuttgart, Madison, Montreal, Pittsburgh, Cleveland, Charlotte, Toronto, Phoenix and Las Vegas in terms of global integration.

Similarly, calculated ratios by categories result in values extremely close to 0, what is translated into a proportionally similar level of interaction of tourists and residents. Only Toronto’s monuments and Charlotte’s mobility infrastructures and services average ratio ranks positive, showing a proportionally higher interaction of tourists than residents. In fact, businesses such as mobility-related infrastructures and services (airports, train, bus and taxi stations, parking and vehicle rental services, among others) together with businesses and venues marked as “Monuments, landmarks & heritage” (churches, cathedrals, castles, architectural tours and historical buildings) show the higher level of integration between tourists and residents. In general, a higher proportion of businesses with positive ratios can also be found among these categories, as represented by Table 3 (page 45):

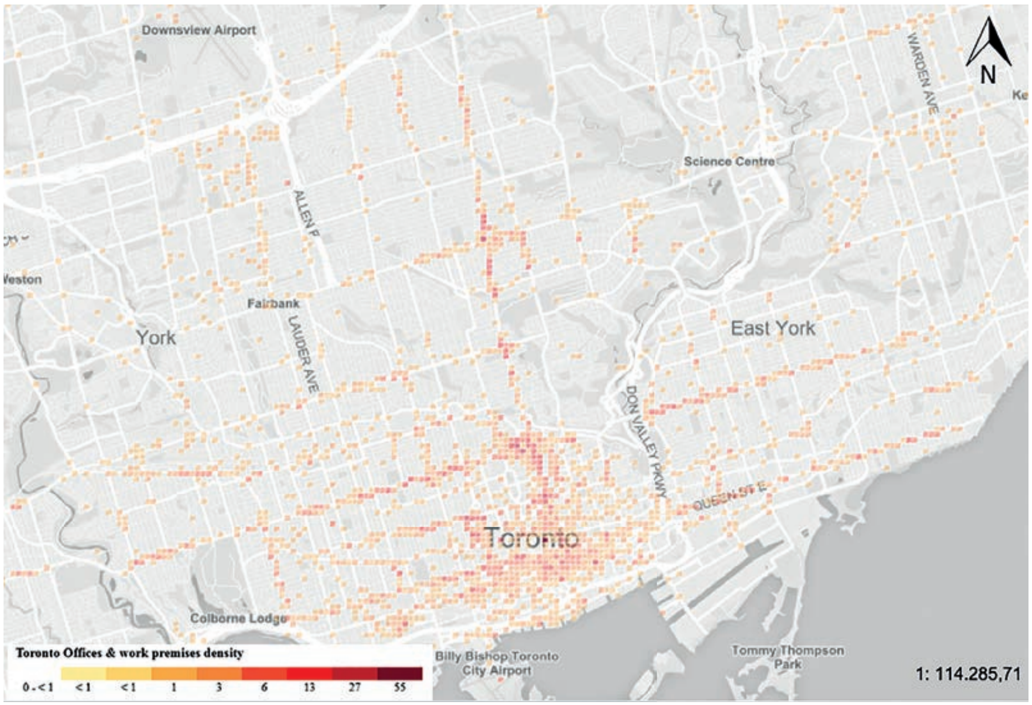
1 All processes and maps have been compiled and are openly accessible at the following GitHub repository: https://oromerob.github.io/spatial_distribution_tourist_perception/

Figure 1 Spatial distribution of Businesses concentration in Toronto (per tile of 640 m²)



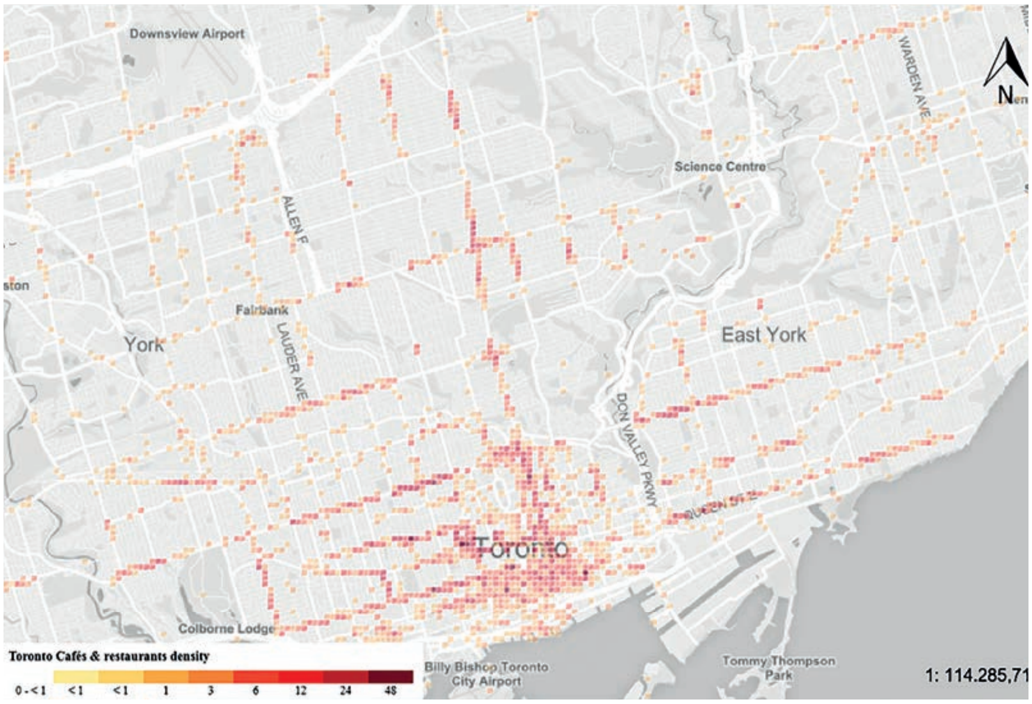
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 2 Spatial distribution of Offices and diverse work premises in Toronto (per tile of 640 m²)



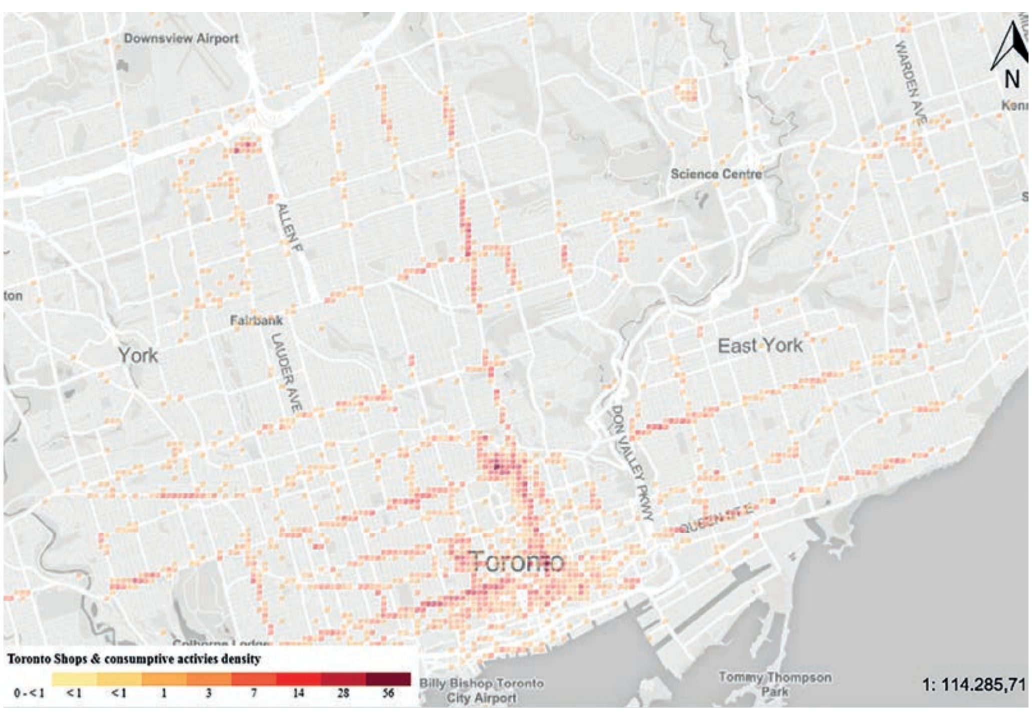
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Fig. 3 Spatial distribution of Cafes, bars, restaurants & catering activities in Toronto (per tile of 640 m²)



Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 4 Spatial distribution of Shops and stores in Toronto (per tile of 640 m²)



Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Table 3 Distribution by metropolitan areas of the proportion of business type showing positive integration ratios

Metropolitan Area	% Offices and diverse work premises	% Shops and stores	% Cafes, bars, restaurants & catering activities	% Nightclubs, bars & nightlife offer	% Sport venues and related services	% Cinemas, concert venues & theatres	% Museums, art galleries	% Public mobility infrastructures & services	% Monuments, landmark & heritage	% Private transport services
Champaign	13,22%	6,41%	2,10%	4,27%	18,75%	19,51%	7,14%	13,04%	25,00%	18,75%
Charlotte	9,40%	4,08%	2,53%	2,96%	3,72%	15,28%	4,63%	16,67%	3,70%	12,99%
Cleveland	11,61%	5,77%	2,20%	2,84%	5,88%	11,28%	10,20%	20,59%	11,11%	10,96%
Edinburgh	4,12%	3,96%	3,41%	3,01%	4,92%	5,26%	1,59%	3,85%	14,29%	0,00%
Las Vegas	4,85%	4,64%	2,62%	7,48%	3,71%	7,67%	8,85%	26,76%	8,06%	28,42%
Madison	6,28%	7,72%	5,93%	6,21%	10,19%	8,33%	6,25%	0,00%	0,00%	11,90%
Montreal	15,86%	4,62%	2,81%	4,51%	2,86%	10,88%	10,34%	9,09%	19,23%	11,76%
Phoenix	7,47%	5,10%	2,45%	3,93%	7,47%	12,84%	8,49%	25,51%	7,65%	15,54%
Pittsburgh	9,49%	4,68%	2,00%	2,91%	7,01%	14,44%	9,28%	24,24%	7,02%	18,97%
Stuttgart	14,12%	4,99%	5,50%	6,23%	8,70%	10,13%	5,88%	42,11%	0,00%	15,38%
Toronto	9,27%	5,47%	2,17%	3,18%	3,79%	7,19%	8,50%	10,56%	17,14%	15,00%

Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017

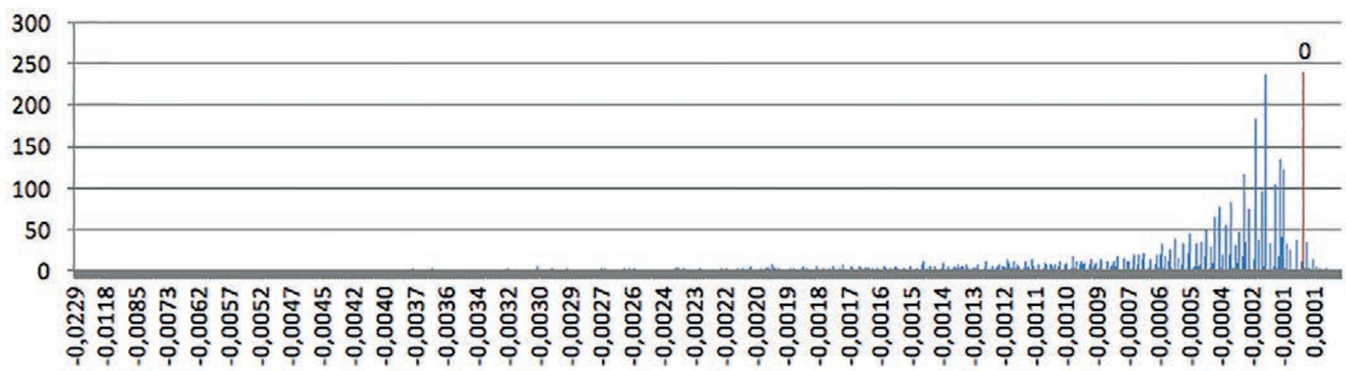
On the contrary "nightclubs, bars & nightlife offer", and "cafes, bars, restaurants & catering activities" are found out to concentrate more venues with a lower level of integration, mostly negative, that being those where a higher proportion of residents than tourists posting reviews in such businesses. Still, the majority of businesses integration ratio values remain very close to 0 in all cases (being 0,000158214689460 the highest, and -0,007237375536728 the lowest). In term of business numbers, these categories also present the lowest proportion of businesses showing positive ratios (an average of 3%), what corroborates the finding of that there are more nightlife-oriented, and catering services-oriented businesses where proportionally more residents than tourists go, and that those frequented by tourists are also frequented by residents in a very similar proportion. This distribution can be clearly seen in Figures 5 and 6 (Cleveland and Pittsburgh respectively) distributions of the number of "cafes, bars, restaurants & catering activities" businesses by integration ratio distribution (next page).

As shown by both figures, there is a higher frequency of venues close to 0 values, and a large dispersion of very few businesses closer to the most negative of the values (far left of the horizontal axe).

Differences are observed when analysing frequency dispersion of more integrated categories such as Phoenix's "Museums, art galleries" businesses group (figure 7, at next page), whose ratio dispersion is much smaller and better balanced between positive and negative values. As illustrated by Phoenix's example below, businesses frequencies are distributed between smaller ratio values. Still, the overall distribution trend identified in figures 5 and 6 is reproduced at a lower scale.

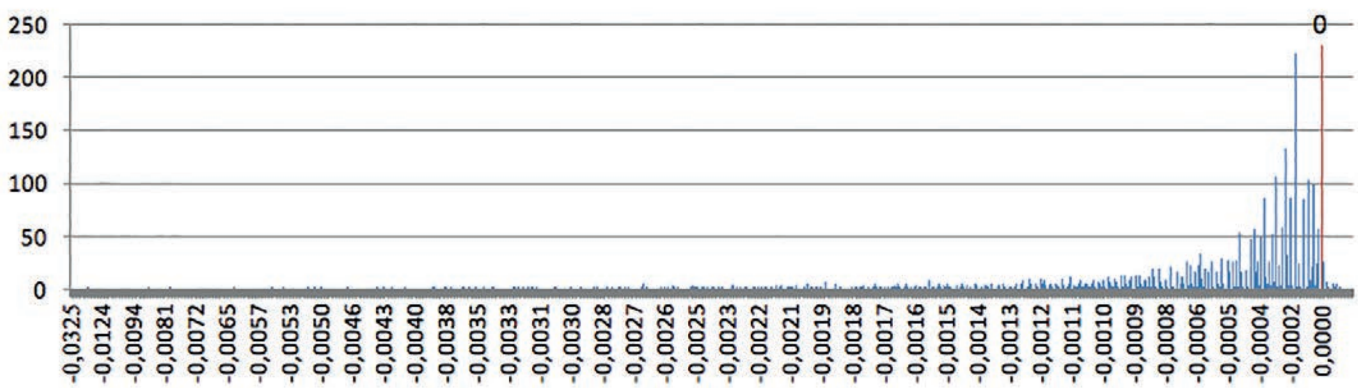
In applying, Pearson's correlation coefficient, it has been found that there is a strong inverse correlation between the number of reviews registered per venue and the integration ratio R. In other words, the increasing number of reviews a business has, the lower ratio (closer to -1) the business will have. This is especially relevant when considering that residents' level of interaction with businesses through social media is assumed to be lower than tourists', as it appears

Figure 5 Cleveland metropolitan area's "cafes, bars, restaurants & catering activities" businesses frequency by integration ratio



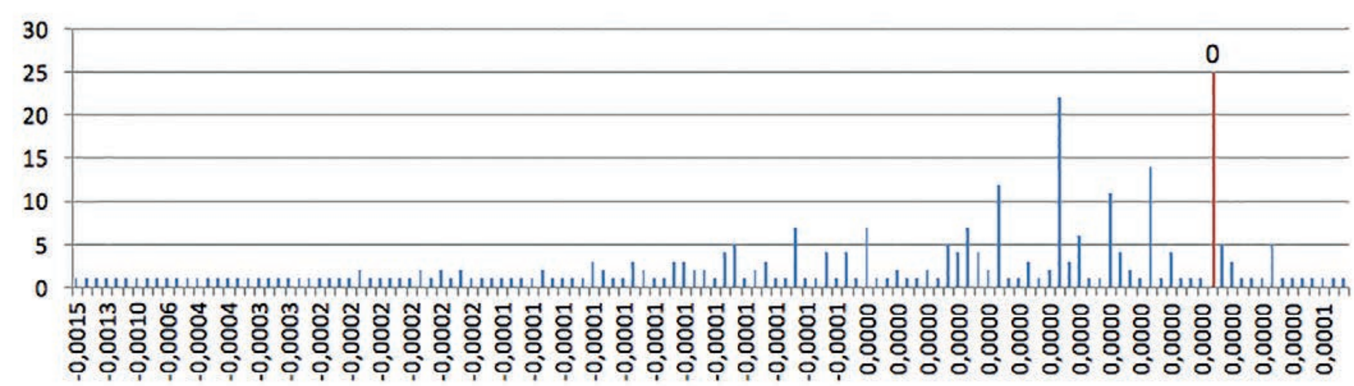
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 6 Pittsburgh metropolitan area's "cafes, bars, restaurants & catering activities" businesses frequency by integration ratio



Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 7 Phoenix metropolitan area's "museums, art galleries" businesses frequency by integration



Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

that residents write more reviews on businesses' profiles where the presence of tourists is lower. Only Madison and Edinburgh are exceptions to this rule, where absolutely no correlation has been found at a city level. The remaining metropolitan areas show diverse results, but "Shops and stores", followed by "Cafes, bars, restaurants & catering activities", and "Nightclubs, bars & nightlife offer", are without any doubt the category where the businesses with more reviews are also the ones proportionally more frequented by residents.

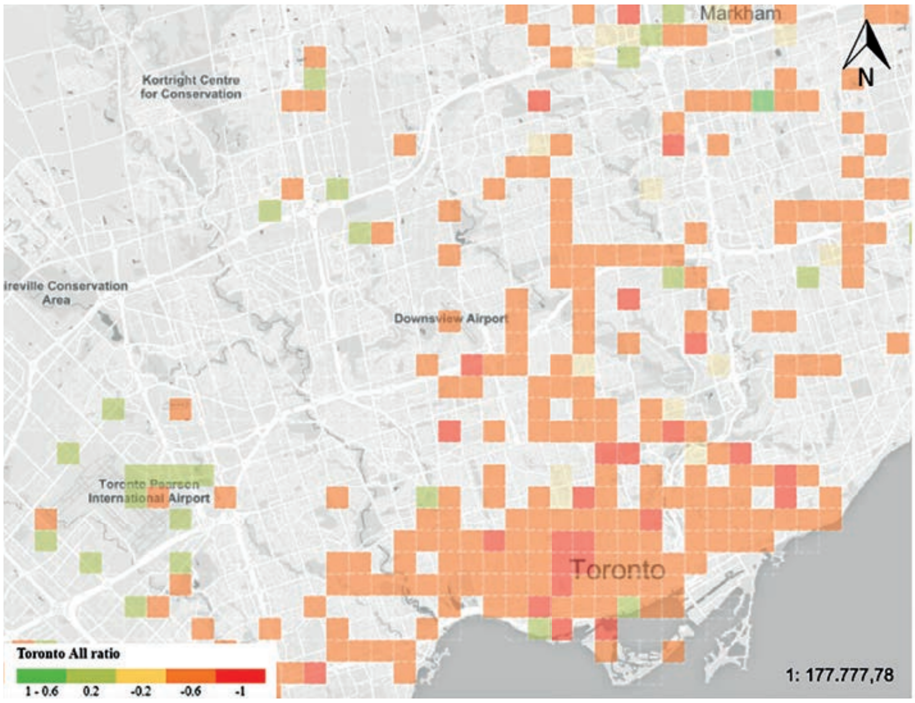
Tourists and residents integration spatial distribution

When displaying Z results at a city level for all categories, metropolitan areas with the higher amount of listed businesses show a clear integration tendency when visualized at tile zoom 15 (409.600 m^2), particularly from the year 2015 as illustrated in figures 8, 9, 10 and 11 (next pages) with Toronto's and Charlotte's examples. On the other side, metropolitan areas with a lower amount of venues such as Montreal, Pittsburgh, Edinburgh and Stuttgart do not follow this trend and show fragmented results depending on the year analysed.

No significant pattern has been identified when analysing temporal evolution of functional category-associated ratios, except for the fact that those areas where airports are located seem to maintain a positive integration ratio. Also, no visible cluster of positive ratio that remains stable or whose extension increases can be identified at tile zoom 15. Results displayed at tile zoom 18 (6.400m^2) are consistent with this and show significant differences among years and metropolitan areas without a clear pattern. The well-known Strip tourist area in Las Vegas is an exception as, as displayed below in figure 12 (next pages), it appears to increasingly concentrate tiles at zoom 18 that show positive ratio.

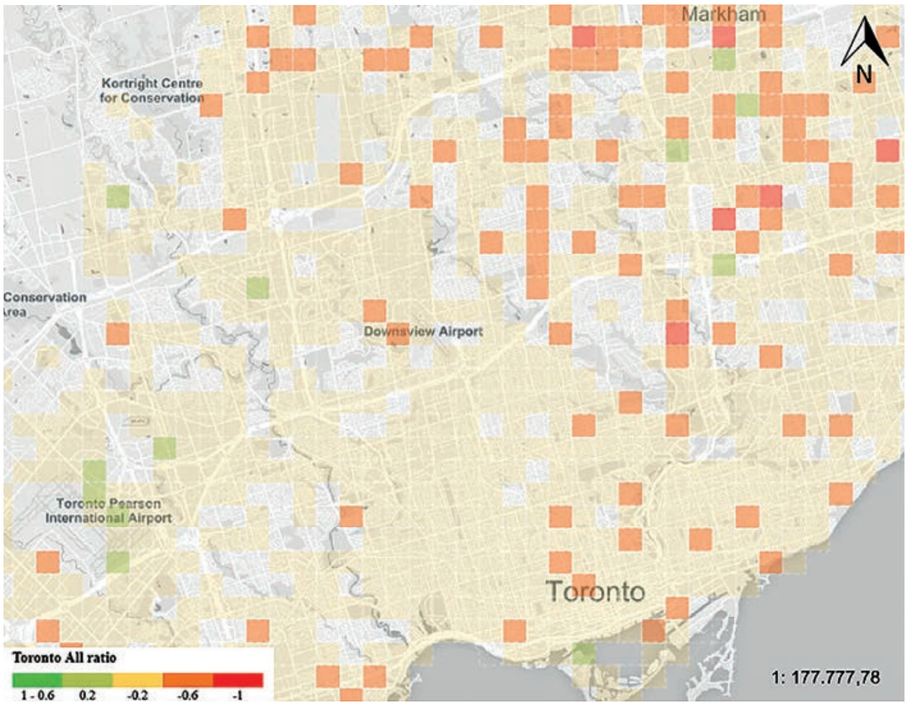
Results suggest that spatial integration for all categories (excepting international airports) increases over time, which implies a growing spatial dispersion of the tourist activity. This is consistent with the reviewed literature, where tourists-oriented spaces are stated to be integrated with the rest of urban functions. Still, it has to be noted that the lack of qualitative data doesn't allow to presuppose any displacement of urban functions, as hinted by Judd's (1999) "tourist bubbles" conceptualization. Also, the differences in data amounts between metropolitan areas and functional categories seem to condition the identification of integration patterns. For this reason, the dataset analysed here could be complemented with additional big data sources oriented to different targets, as well as official tourist affluence statistics to reduce possible bias and contextualized the results obtained.

Figure 8 Toronto metropolitan area's 2008 integration ratio (per tile of 409.600 m^2)



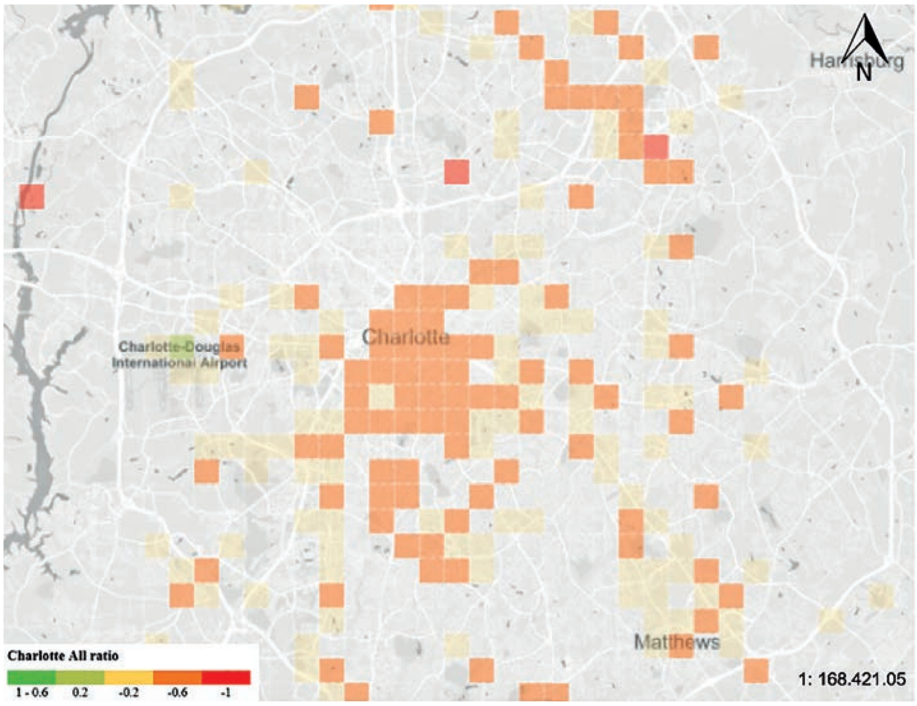
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 9 Toronto metropolitan area's 2017 integration ratio (per tile of 409.600 m²)



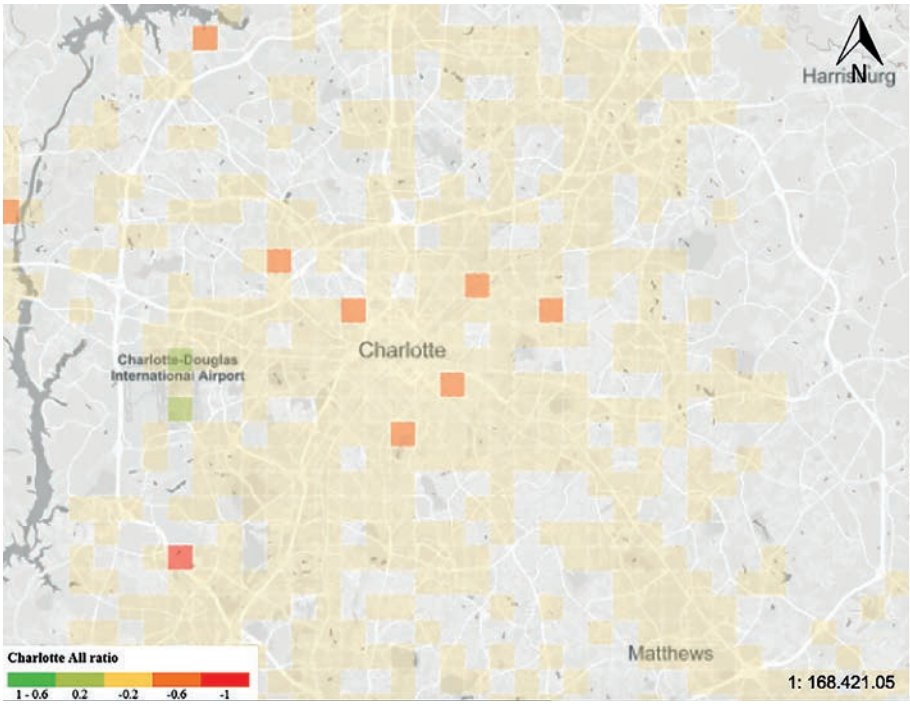
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 10 Charlotte metropolitan area's 2008 integration ratio (per tile of 409.600 m²)



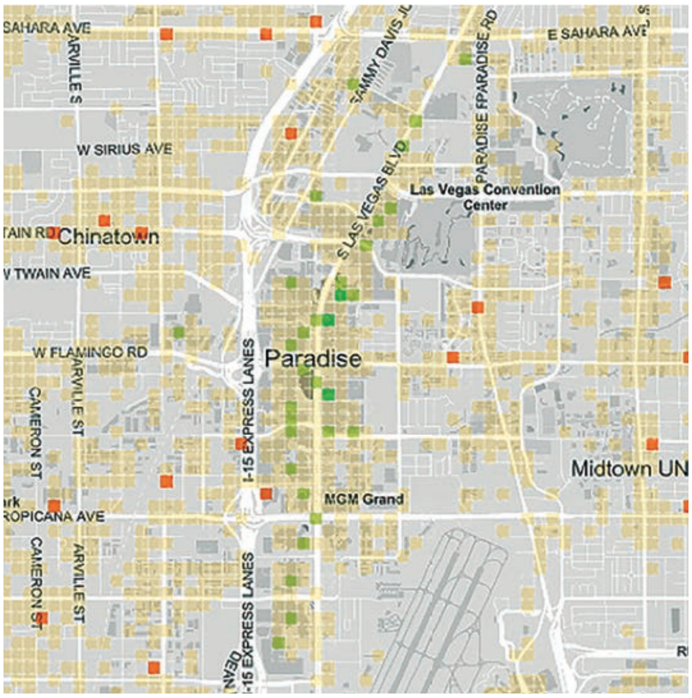
Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 11 Charlotte metropolitan area's 2017 integration ratio (per tile of 409.600 m^2)



Source: Own elaboration based on Yelp's 11th Round Challenge dataset and comprising data from 2004 to 2017.

Figure 12 Las Vegas metropolitan area's 2017 integration ratio



Source: Own elaboration.

Conclusions

The initial data exploration presented in this paper sought to analyse the level of integration between tourists and residents in several metropolitan areas. The methodology developed covers different phases that go from the initial geographical clustering of venues, to the identification of user types using a specially crafted algorithm based on several previous studies, and to the calculation of integration ratios, and that ultimately constitutes the most significant contribution to the research field. In this case, LBSNs have proven to be a source of a significant amount of data that, if used complementary with ground-based knowledge, give valuable knowledge about the urban tourism phenomenon. Though, the lack of structured data and the high amount of information require very specific methods to be developed ad-hoc for each different type of analysis.

Results were expected to show a similar functional integration level between tourists and residents as previous literature states that both user groups make a similar usage of the urban space, even if tourists confront specific constraints related to their short length of stay and limited budget. Although the obtained results appear to confirm that tourists and residents interact with the same type of business, further detail could be beneficial to avoid the loss of nuances in the quantitative treatment of text variables meant to be complementary, as it happens with the different tags used in categorising businesses.

Additionally, results were also expected to show a higher concentration of the proportion of tourists in a less scattered area, as previous studies lead one to think. Despite this, tourist activity seems to increasingly spread in the urban space over time, without concentrating enough in specific areas to result in the loss of integration between tourists and residents. However, it can be argued that the source in which the research is based lead to bias due to the own preferences of its users who, as in any other LBSN, are also the content generators. Specifically, YELP promotes itself as being especially popular among locals, a premise consistent with the obtained results. Future research could contribute to overcoming this limitation by introducing alternative data sources to obtain additional data and different user profile types. Furthermore, it can also be discussed whether the dismissal of such an important amount of users has significantly affected the obtained results, and future research could be worthwhile to sharpen the first step of the algorithm presented here. Nevertheless, this initial exploration clearly allows the identification of delimitated areas that present a higher business concentration with a clear associated multifunctionality, in line with previous studies that outline this particularity of urban tourist destinations. It is therefore recommended to provide continuity to studies that incorporate big data as information sources, as they have proven to be able to provide new information that, in combination with others, can support tourism planning and management decisions.

References

Arizona Office of Tourism. (2016). *Phoenix & Central Region Visitor Profile report*. Retrieved from http://www.stat.gouv.qc.ca/quebec-chiffre-main/pdf/qcm2017_fr.pdf

Ashworth, G., & Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. *Tourism Management*, 32(1), 1–15. https://doi.org/10.1016/j.tourman.2010.02.002

Batista e Silva, F., Marín Herrera, M. A., Rosina, K., Ribeiro Barranco, R., Freire, S., & Schiavina, M. (2018). Ana–lysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, 101–115. https://doi.org/10.1016/j.tourman.2018.02.020

Brandt, T., Bendler, J., & Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information and Management*, 54(6), 703–713. https://doi.org/10.1016/j.im.2017.01.004

Burtenshaw, D., Bateman, M., & Ashworth, G. (1991). *The European city. A western perspective*. (2nd ed.). London: David Fulton Publishers Ltd.

Edwards, D., Griffin, T., & Hayllar, B. (2008). Urban Tourism Research. Developing an Agenda. *Annals of Tourism Research*, 35(4), 1032–1052. https://doi.org/doi:10.1016/j.annals.2008.09.002

Fernández Güell, J. M., & López, J. G. (2016). Cities futures. A critical assessment of how future studies are applied to cities. *Foresight*, 18(5), 454–468. https://doi.org/10.1108/FS-06-2015-0032

Florido Trujillo, G., Garzón García, R., & Ramírez López, M. L. (2018). En torno al concepto de sostenibilidad y su compleja aplicación al turismo. el caso del turismo urbano cultural. *International Journal of Scientific Management and Tourism* (2018), 269–302.

Füller, H., & Michel, B. (2014). "Stop Being a Tourist!" New Dynamics of Urban Tourism in Berlin-Kreuzberg. *International Journal of Urban and Regional Research*, 38(4), 1304–1318. https://doi.org/10.1111/1468-2427.12124

García-Hernández, M., de la Calle-Vaquero, M., & Yubero, C. (2017). Cultural heritage and urban tourism: Historic city centres under pressure. Sustainability (Switzerland), 9(8). https://doi.org/10.3390/su9081346

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002

Getz, D. (1993a). Planning for tourism business districts. *Annals of Tourism Research*, 20(3), 583–600. https://doi.org/10.1016/0160-7383(93)90011-Q

Getz, D. (1993b). Tourist shopping villages. Development and planning strategies. *Tourism Management*, 14(1), 15–26. https://doi.org/10.1016/0261-5177(93)90078-Y

Hayllar, B., & Griffin, T. (2005). The precinct experience: A phenomenological approach. *Tourism Management*, 26(4), 517–528. https://doi.org/10.1016/j.tourman.2004.03.011

Hayllar, B., Griffin, T., & Edwards, D. (2008). *City Spaces-Tourist Places: Urban tourism precincts*. (B. Hayllar, T. Griffin, & D. Edwards, Eds.), City Spaces–Tourist Places. Butterworth-Heinemann. https://doi.org/10.4324/9780080878270

Jansen-Verbeke, M. (1986). Inner-city tourism: resources, tourists and promoters. A*nnals of Tourism Research*, 13(1), 79–100. https://doi.org/https://doi.org/10.1016/0160-7383(86)90058-7

Jansen-Verbeke, M. (1998). Tourismification of Historical Cities. *Annals of Tourism Research*, 25(4), 739–742. https://doi.org/10.1016/S0160-7383(98)00015-2

Jansen-Verbeke, M., & Ashworth, G. (1990). Environmental integration of recreation and tourism. Annals of Tourism Research, 17(4), 618–622. https://doi.org/10.1016/0160-7383(90)90034-O

Judd, D. R. (1999). *The Tourist City*. (D. R. Judd & S. S. Fainstein, Eds.). New Haven and London: Yale University Press.

Kádár, B. (2013). Differences in the spatial patterns of urban tourism in Vienna and Prague. *Urbani Izziv*, 24(2), 96–111. https://doi.org/10.5379/urbani-izziv-en-2013-24-02-002

Kannisto, P. (2018). Travelling like locals: Market resistance in long-term travel. *Tourism Management*, 67, 297–306. https://doi.org/10.1016/j.tourman.2018.02.009

Kuo, C.-L., Chan, T.-C., Fan, I.-C., & Zipf, A. (2018). Efficient Method for POI/ROI Discovery Using Flickr Geotagged Photos. *ISPRS International Journal of Geo-Information*, 7(3), 121. https://doi.org/10.3390/ijgi7030121

Leung, R., Vu, H. Q., & Rong, J. (2017). Understanding tourists' photo sharing and visit pattern at non-first tier attractions via geotagged photos. *Information Technology and Tourism*, 17(1), 55–74. https://doi.org/10.1007/s40558-017-0078-3

Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research*, 33(2), 403–423. https://doi.org/10.1016/j.annals.2005.12.002

Li, D., Zhou, X., & Wang, M. (2018). Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities*, 74(January), 249–258. https://doi.org/10.1016/j.cities.2017.12.012

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

MacCannell, D. (1976/2017). *El Turista: una nueva teoría de la clase ociosa* [The Tourist: a New Theory of the Leisure Class]. Editorial Melusina.

Maeda, T., Yoshida, M., Toriumi, F., & Ohashi, H. (2018). Extraction of Tourist Destinations and Comparative Analysis of Preferences Between Foreign Tourists and Domestic Tourists on the Basis of Geotagged Social Media Data. *ISPRS International Journal of Geo-Information*, 7(3), 99. https://doi.org/10.3390/ijgi7030099

Marine-Roig, E., & Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162–172. https://doi.org/10.1016/j.jdmm.2015.06.004

Mukhina, K. D., Rakitin, S. V., & Visheratin, A. A. (2017). Detection of tourists attraction points using Instagram profiles. *Procedia Computer Science*, 108(C), 2378–2382. https://doi.org/10.1016/j.procs.2017.05.131

Önder, I. (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives*, 21, 54–58. https://doi.org/10.1016/j.tmp.2016.11.002

Page, S. J. (1995). *Urban Tourism*. London: Routledge.

Page, S. J., & Hall, C. M. (2003). *Managing Urban Tourism*. Harlow: Prentice Hall.

Pearce, D. G. (1998). Tourist districts in Paris: Structure and functions. *Tourism Management*, 19(1), 49–65. https://doi.org/10.1016/S0261-5177(97)00095-2

Pearce, D. G. (2001). An integrative framework for urban tourism research. *Annals of Tourism Research*, 28(4), 926–946. https://doi.org/10.1016/S0160-7383(00)00082-7

Pranata, I., & Susilo, W. (2016). Are the most popular users always trustworthy? The case of Yelp. *Electronic Commerce Research and Applications*, 20, 30–41. https://doi.org/10.1016/j.elerap.2016.09.005

Rogerson, C. M., & Rogerson, J. M. (2016). Intra-urban spatial differentiation of tourism: Evidence from Johannesburg, South Africa. *Urbani Izziv*, 27(2), 125–137. https://doi.org/10.5379/urbani-izziv-en-2016-27-02-004

Sakoda, J. M. (1981). A Generalized Index of Dissimilarity. *Demography*, 18(2), 245–250. https://doi.org/10.2307/2061096

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13–25. https://doi.org/10.1016/j.tourman.2017.11.001

Sapountzi, A., & Psannis, K. E. (2016). Social networking data analysis tools & challenges. *Future Generation Computer Systems,* 86, 893–913. https://doi.org/10.1016/j.future.2016.10.019

Scherrer, L., Tomko, M., Ranacher, P., & Weibel, R. (2018). Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Science*, 1–21. https://doi.org/10.1140/epjds/s13688-018-0147-7

Shao, H., Zhang, Y., & Li, W. (2017). Extraction and analysis of city's tourism districts based on social media data. *Computers, Environment and Urban Systems*, 65, 66–78. https://doi.org/10.1016/j.compenvurbsys.2017.04.010

Shoval, N., & Raveh, A. (2004). Categorization of tourist attractions and the modeling of tourist cities: Based on the co-plot method of multivariate analysis. *Tourism Management*, 25(6), 741–750. https://doi.org/10.1016/j.tourman.2003.09.005

Stock, K. (2018). Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*, (May), 1–32. https://doi.org/10.1016/j.compenvurbsys.2018.05.007

Vera Rebollo, J. F., López Palomeque, F., Marchena Gómez, M. J., & Anton Clavé, S. (2011). *Análisis Territorial del Turismo y Planificación de Destinos Turísticos*. (J. F. Vera Rebollo, Ed.). Valencia: Tirant lo Blanch.

Zhou, B., Tang, X., Zhang, H., & Wang, X. (2014). Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8), 1586–1599. https://doi.org/10.1109/TPAMI.2014.2300484

Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems*, 54, 144–153. https://doi.org/10.1016/j.compenvurbsys.2015.07.006

Author's Details

Aurélie Cerdan Schwitzguébel
aurelie.cerdan@cett.cat

Oriol Romero Bartomeus
oriol.bartomeus@clarivate.com