Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy

(Article begins on next page)

Elsevier Editorial System(tm) for Tourism Management

Manuscript Draft

Manuscript Number: JTMA-D-15-01046R2

Title: Mapping Cilento: Using Geotagged Social Media Data to Characterise Tourist Flows in Southern Italy

Article Type: Case Study

Keywords: Data Mining; Visual Analytics; Flow Analysis; Geotagged Social Media Data

Corresponding Author: Mr. Alvin Chua,

Corresponding Author's Institution: KU Leuven

First Author: Alvin Chua

Order of Authors: Alvin Chua; Loris Servillo; Ernesto Marcheggiani; Andrew Vande Moere

Abstract: New sources of geotagged information derived from social media like Twitter show great promise for geographic research in tourism. This paper describes an approach to analyze geotagged social media data from Twitter to characterize spatial, temporal and demographic features of tourist flows in Cilento - a regional tourist attraction in southern Italy. It demonstrates how the analysis of geotagged social media data yields more detailed spatial, temporal and demographic information of tourist movements in comparison to the current understanding of tourist flow patterns in the region. The insights obtained from our case study illustrate the potential of the proposed methodology yet attention should be paid to biases in the data as well as methodological limitations when drawing conclusions from analytical results.

Response to Reviewers

We would like to thank the reviewers for the valuable feedback on our work. We agree with most of the comments and attempt to address each to the best of our abilities. We hope that the revisions are able to the meet the reviewers' expectations. In addition to the key points stated below, we have corrected grammatical and spelling errors. A substantial amount of effort has also gone into verifying that the references are consistently formatted and that missing information like the date, author names and page numbers are resolved. For the following points, we will use p. x, l. y (page x, line y) to indicate where revisions have been made.

1. **Is the proposed tourist flow analysis approach a better method than previous research works dealing with the same research topic?**

   We do not claim that our work is an improvement on, or replacement for, other approaches to characterize tourist flows. However, it does present a substantial advancement in terms of detail when compared to previous attempts to do so with GSMD. The aforementioned points are located between p. 30, l. 581 and p. 31, l. 586 as well as p. 32, l. 629 and p. 32, l. 632.

2. **How can better performance of flow analysis be assured when analytical approach goes beyond the case region and is faced with data biases as well as methodological limitations?**

   We acknowledge the limited discussion in the previous submission. Thus we have elaborated on the subject between p. 32, l. 632 and p. 33, l. 640.

3. **It will be better to provide a comprehensive literature review on characterizing tourist flows using GSMD (or specifically geotagged Twitter data) to positioning this paper in the targeted research field, and show the significance of contribution in both the academic and practical domains.**

   We appreciate the constructive comments on this issue and have further elaborated on the subject between p. 11, l. 242 and p. 12, l. 263. The change includes new 10 references on the use of GSMD in analysis of tourist travel behaviour. Yet only 4 of 10 of these references, to the best of our knowledge, have attempted to characterize certain features of tourist flows.

4. **It will be helpful to reorganize section 5 and provide a subsection to describe the public/business implications of the proposed approach in order to reveal its practical importance and guide the way for future adoption.**

   We agree with the comment and describe possible scenarios for public administrations as well as large and small organizations between p. 33, l. 640 and p. 33, l. 646.

5. **How the flow map optimization is achieved and why the shortest path representation is appropriate require explanation.**

   We have distinguished between the description of the flow map optimization technique in Section 4.3.1 and its implementation in Section 4.3.2 to maintain a certain degree of structure in the paper. Indeed, we have overlooked the rationale for our implementation of the shortest path representation. Thus, this is added to the text at p. 18, l. 390.

***Title page with author details**

Mapping Cilento: Using Geotagged Social Media Data to Characterize Tourist Flows in Southern Italy

Alvin Chua[1], Loris Servillo[1], Ernesto Marcheggiani[23] and Andrew Vande Moere[345]

[1] (alvin.chua, loris.servillo, andrew.vandemoere)@asro.kuleuven.be
[2] e.marcheggiani@univpm.it

[3] Department of Architecture, KU Leuven, Belgium
[4] Department of Earth and Environmental Sciences, KU Leuven, Belgium
[5] Department D3A, University of Marche, Italy

***Highlights**

- We introduce a novel approach to tourist flow analysis based on geotagged social media data.
- Our approach is structured around three research questions that investigate the spatial, temporal and demographic features of tourist flows.
- Our analysis yields more detailed spatial, temporal and demographic information of tourist movements in comparison to the current understanding of tourist flow patterns in the region.

Mapping Cilento: Using Geotagged Social Media Data to Characterize Tourist Flows in Southern Italy

Abstract

New sources of geotagged information derived from social media like Twitter show great promise for geographic research in tourism. This paper describes an approach to analyze geotagged social media data from Twitter to characterize spatial, temporal and demographic features of tourist flows in Cilento - a regional tourist attraction in southern Italy. It demonstrates how the analysis of geotagged social media data yields more detailed spatial, temporal and demographic information of tourist movements in comparison to the current understanding of tourist flow patterns in the region. The insights obtained from our case study illustrate the potential of the proposed methodology yet attention should be paid to biases in the data as well as methodological limitations when drawing conclusions from analytical results.

Keywords: *Data Mining; Visual Analytics; Flow Analysis; Geotagged Social Media Data;*

1. Introduction

Flow analysis is a topic of theoretical interest and practical importance in many disciplines. “Flow” commonly refers to the collective movement of people or other abstract concepts like energy, material and information, from a particular location to another. Flow analysis is conventionally conducted to study spatial dynamics and understand how the environment influences the way people move. For instance, interest in modeling traffic flows emerged from the need to identify factors that lead to congestion on transportation infrastructures (Nagatani, 2002). Likewise, insight into routine flow patterns such as journeys between home and work is crucial for the conceptualization of functional urban areas (Sykora & Mulicek, 2014), urban hierarchies (Christaller, 1964) and other territorial structures for policy enactment.

Tourism plays a major role in many regional economies (Ashley, De Brine, Lehr, & Wilde, 2007) and accounts for a substantial amount of human movement (Schlich & Axhausen, 2003). To meet the planning demands of the tourism industry, it has become increasingly important to monitor and analyze the flows of tourists (Williams, 1998). Access to detailed records of travel routes enables the design of policies that prevent capacity overload on the transportation infrastructure and resolve travel barriers between tourist destinations (Prideaux, 2000). Similarly, attractions can be improved or developed according to the preference of different tourist demographics (Lew & McKercher, 2006). While techniques to study routine travel habits are well established, relatively fewer methods have been developed to support the analysis of tourist flows, which are comparably sporadic. One reason for the lack of progress thus far stems from difficulties with data acquisition as tourism often involves travel between urban and rural spaces (Christaller, 1964; Mansfeld, 1990). From a sensing perspective, this poses peculiar challenges with spatial-temporal precision as well as cost in resources (See Section 3).

Over the past decade, large geotagged datasets have become increasingly commonplace due to the proliferation of sensor networks and portable devices like smartphones.

Termed “Big Data” due to the sheer volume of records that emerge from real-time sensing (Kitchin, 2014), such datasets typically contain information of activities or processes linked to the space and time where they occur. In the domain of “Smart City” research (Kitchin, 2014), much has been accomplished with the use of “Big Data” to study human movement. Smart card data from subway (Roth, Kang, Batty, & Barthélemy, 2011) and bike sharing systems (Beecham, Wood, & Bowerman, 2014), taxi journey GPS logs (Ferreira, Poco, Vo, Freire, & Silva, 2013) as well as cellular call data records (Sevtsuk & Ratti, 2009) have provided new opportunities to develop greater understanding of mobility patterns in urban environments (Batty, et al., 2012). In recent years, social media has exercised a powerful influence on the tourism industry as people increasingly rely on virtual communities, personal blogs and networks like Flickr, Twitter and Instagram for travel information (Xiang & Gretzel, 2010). Geotagged Twitter data in particular contains larges amounts of up-to-date content for most locations worldwide (Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013). From this perspective, the constant availability of highly granular user-generated data serves as a valuable source of information to study the movements of tourist as well as to understand their travel preferences.

In this paper, we describe the use of geotagged social media data to characterize the spatial, temporal and demographic features of tourist flows. Based on a case study situated in Cilento - a tourist venue in southern Italy, we will demonstrate how our analytical approach, operationalized with geotagged Twitter data, addresses the challenge of tracking large numbers of tourists across a large region. More importantly, we will show how the insights we acquired provide more spatial detail than the current understanding of tourist movements in the case study context, prompting a discussion on the value of our approach in contrast to the methods previously utilized to analyze tourist flows in Cilento.

The paper is organized in the following manner. First, we outline the case study context and describe the limitations faced by existing data of tourist movements in Cilento. Next we present different methods to gather data of tourist flows, elaborating on the advantages and limitations of each method to determine a suitable alternative. Thereafter, we explain our analytical methodology and technical details related to data processing and visualization. This is followed by a report and discussion of our findings structured around three research questions. Finally, we compare the insights we obtained to existing knowledge of tourist movements in the case study context, and discuss the relative merits of our approach.

2. Case Study

Cilento is a well-known tourist venue located in southern Italy where, for the last two years, policy makers have engaged in a national interest project[6] funded by European and state agencies[7] to foster the exchange of best practices in sustainable tourism between developed and under-developed regions in Italy. In this particular context, the objective is to develop a local strategy for tourism that encourages economic development and territorial cohesion. The Cilento region comprises of 31 municipalities, spanning approximately 490,000 hectares. The landscape comprises of different environments including a picturesque coastline and mountainous inland dotted with multiple UNESCO heritage sites (e.g. Paestum, Punta Licosa, Capo Palinuro).

[6] TOOKMC: Transfer Of Organized Knowledge Marche-Cilento

[7] Italian Ministry of Economic Development and Department for Development and Territorial Cohesion

While majority of the settlements are located close to the coast, the inner boundaries of Cilento mainly consist of land parcels for agriculture and nature conservation. Tourism represents a significant fraction of the regional economy. In 2015, an estimated 9% of the regional GDP was derived from tourism and that 11% of the workforce was employed to staff the sector (WTTC, 2015). As of late, the region has suffered from low economic performance, due to changing tourist demographics and spending power. In light of this circumstance, a comprehensive revision to existing tourism policies is required to reverse negative economic trends.

We conducted a workshop in collaboration with local policy makers and experts from other institutions to familiarize with the region. This workshop spanned the duration of a week involving direct observation at several coastal and inland attractions followed by a seminar where participants described methods for monitoring tourist activity on the ground. The insights obtained from this workshop revealed a disproportionate distribution of tourist activity at the coast as opposed to the interior. Contrary to prevailing policies that promote the region to other European communities, young urbanites from major Italian cities are observed to account for the bulk of tourism. Members of this demographic group prefer to engage in beach activities around coastal resorts instead of visiting natural or heritage attractions situated inland. Other tourists have expressed interest in these attractions but the absence of public transportation discourages prospective travel. The result is a polarized distribution of economic activity in the region, and the potential loss of jobs that rely on the inland economy.

2.1. Research Questions and Data Criteria

While existing information is sufficient for a broad understanding of how changes affect the tourism industry, detailed knowledge of tourist flows is required to inform the design of a suitable policy response. There are three specific research questions (RQ) to be addressed:

RQ1. What are the meaningful tourist profiles in the region?

RQ2. What are the valuable patterns of tourists flows in the region?

RQ3. Where are the tourist attractions in the region and how do they differ?

Each RQ investigated a particular set of features related to tourist flows. RQ1 focused on the demographic composition of tourists in order to determine how the movements of various demographic groups differ spatially and temporally from one another. RQ2 investigated spatial and temporal patterns in the data for insight into tourist movements. RQ3 evaluated the relative importance of locations as centers of tourist activity to discover the factors that differentiate well-known attractions from those that are under utilized. Table 1 summarizes the expected outcome from our analysis of tourist flows in the region. The type of information as well as the corresponding level of detail for each feature listed, served as a criteria to determine the quality of data we required.

Table 1. Information expected from the analysis of tourist flows in Cilento.

Feature	Expected Detail (Granularity / Segmentation)	Expected Type of Information
Demographic	Country	x Distinct demographic groups by location of origin. x Distribution of tourists among the demographic groups.
Temporal	Day	x Evolution of tourist numbers in relation to demographic groups. x Identify seasonal peaks in tourist activity.

Spatial	Neighborhood	x Spatial diffusion of different demographic groups within the region. x Describe how each demographic group moves in relation to available transportation. x Identify tourist hubs based on location popularity. x Spatial characteristics of tourist hubs in relation to transportation. x Effect of distance on movements and destination choices.

2.2. Limitations with Existing Data

Till date, data of tourist movement is scarce and disparate due to the awkward administrative status of the region. Unlike formal administrative bodies that work with standardized protocols, member municipalities operate independently from one another. Current practices include a wide spectrum of methods ranging from aggregated tourism statistics at regional or provincial level to other proxies like records of return tourism, surveys at local tourism offices, check-ins to guest accommodations or ticket sales at popular attractions that provide coarse approximations of actual tourist numbers. A summary of the methods is outlined in Table 2. This information is extracted from a survey on the current practice of each municipal tabulated in Appendix 1.

Table 2. Data currently in use for analysis of tourist flows in Cilento.

Method	Sample	Granularity		Demographic Segmentation	Num. of Municipals
		Spatial	Temporal		
Regional Tourism Statistics	From, To, Date, Number of Tourist *Tuscany, Campania, July, 8,041*	Province	Month	Country	7
Regional Transport Statistics	From, To, Date, Number of People, Purpose of Travel *Naples, Sarleno, July, 4,139, Leisure*	Province	Month	Country	2
Rail Travel Statistics	From, To, Date, Number of People *Sorrento, Padula, July, 43, 379*	Municipal	Month	None	1
Provincial Tourism Statistics	From, To, Number of Tourist *Capaccio, Agropoli, July, 7,964*	Municipal	Month	Country	7
Tourism Satisfaction Survey	From, Till, Destinations, Experience (1-5), Country of Origin, Mode of Transport, Purpose of Travel *21/7/14, 4/8/14, [Ascea, Palinuro, Camerota, Castellabate], [4, 4, 4,*	Municipal	Day	Country	5
	4], IT, Rail, Leisure				

Return Tourism Statistics	Address, From, Till, Country of Employment *Via Colombo Cristoforo 23, 02-072014, 05-08-2014, DE*	Street	Day	Country	2
Check-in to Guest Accommodation	Address, From, Till, Country of Origin *B&B Villa Degli Aranci, 16-07-2014, 18-07-2014, NL*	Street	Day	Country	9
Ticket Sales at Attraction	Location, Date, Number of Tourist *Diocesan Museum, 21-07-2014, 12*	Street	Day	None	9
Direct Observation in the Field	Location, Date, Number of Tourist *"Via Bolivar Strada exit", 22-082014, 28*	Irregular	Day	None	6
Reports, Publications & Periodicals	Location, Date, Number of Tourists *Vallo di Diano, 3/8/2013, 412*	Irregular	Irregular	Irregular	7
Anecdotes from Local Staff / Stakeholder	Location, Date, Comment *Ponte Piaggine, 09-08-2014, "Group of four interested in Nature with plans to visit Monte Motola"*	Irregular	Irregular	Irregular	4

As it stands, the quality of existing data is insufficient for our analysis since measurements made at different spatial-temporal granularities or demographic segmentations cannot be jointly analyzed without loss of precision. Regional or provincial tourism statistics in particular, do not provide enough spatial detail because municipals, rather than individual attractions, are specified as intended destinations. Similarly, general keywords like "leisure" or "business" is often defined as the purpose of travel instead of specific terms like "social visit", "sight seeing" or "shopping" which may provide clues to where and why a location is chosen. Another shortcoming of these statistics is that they are compiled irregularly, and may not provide timely information of the activities on the ground. Other proxies like checkins to guest accommodations or ticket sales at popular attractions offer fine-grained measures of tourist numbers at specific locations and regular intervals, but do not indicate how tourist travel between the tracked locations or attempt to divide tourists into distinct demographic segments. In some cases, information is extracted from official reports, publications or periodicals that describe tourist activities. Likewise, information might be obtained from anecdotes of local staff or stakeholders. As such information is provided without specific reference to a geographic coordinate, spatial granularity remains relatively coarse since the precision of post-hoc mapping is limited to an approximated area. Though these methods are in place, tourist activity is unmonitored in some municipals due to the lack of resources.

3. Alternative Methods to Collect Tourist Flow Data

Trajectories are the atomic unit in flow analysis that captures the movement of a subject over space and time. In general, methods to collect trajectories can be broadly classified as observational or non-observational (Shoval & Isaacson, 2009). Observational methods involve tracking a subject by means of direct surveillance or remote sensing, while non-observational methods rely on self-reported information to recount the subject's sequence of movements.

3.1. Non-Observational Methods

Non-observational methods like recall diaries (Cooper, 1981) or self-administered diaries (Thornton, Williams, & Shaw, 1997) are commonly employed in mobility studies as they do not specifically require the use of sensors and can be tuned to obtain certain information that cannot be explicitly tracked, such as the mode of travel or the purpose of a trip. Nonetheless, non-observational methods face several shortcomings since the quality of the data gathered depends on the subjects' efforts and collaboration. In particular, the spatial-temporal precision of the gathered data tends to be comparatively lower than those obtained with observational methods due to the analogue data collection procedure. Furthermore, other issues like high operating cost limit the regularity in which updates can be made as well as the number of subjects that can be tracked. GPS devices can be introduced to compliment the information reported by subjects and is shown to produce more detailed records of movement in several studies of travel habits (Bakillah, et al., 2014; Mavoa, Oliver, Witten, & Badland, 2011; Neuhaus, 2010), yet the financial cost of procuring GPS devices and the time required to instruct subjects on their use continue to impede large-scale adoption of this technology for research purposes. GPS trajectories can also be obtained from volunteers who share their data publically online (Sieber, 2006) but the number of subjects who participate and the locations that they are active in are relatively limited.

3.2. Observational Methods

Apart from direct surveillance of subjects in the field, most observational methods rely on sensors to passively track movement. The primary benefit of sensors is the ability to continuously collect data throughout the monitoring period. While the type of sensors deployed for observation may vary, “location” and “timestamp” are standard measurements that are often provided as output. Over the past decade, data from an extensive range of sensors has been in use, but cellular call data records (CDR) and geotagged data from social networks like Twitter, Foursquare or Flickr have emerged as the primary source of trajectory information.

CDR contains information of cell phone activities generated by cellular base stations. To analyze movement in CDR, space-partitioning techniques are first employed to divide a given territory into subspaces determined by the locations of cellular base stations. The position of a cell phone is then approximated to the location of the particular base station responsible for routing its signal at a given moment in time. An estimated trajectory is finally derived by chronologically ordering the locations of the base stations that served a particular cell phone. While analysis of CDR has already led to important discoveries of human mobility habits (Gonzalez, Hidalgo, & Barabasi, 2008; Song, Qu, Blumm, & Barabási, 2010), and applied in studies of tourist flows

(Girardin, Calabrese, Fiore, Ratti, & Blat, 2008), its proprietary value make obtaining such data challenging. More importantly, the spatial granularity of CDR tends to be coarse in rural environments due to sparse distribution of cellular infrastructure (Shoval & Isaacson, 2009). CDR is also often anonymized for privacy preservation thus the demographic information of the tracked subjects cannot be explicitly extracted.

Geotagged social media data (GSMD) is considered a valuable proxy of human movement (Hawelka, et al., 2014) and was the focus of several important studies concerned with how factors like social-economic status (Cho, Myers, & Leskovec, 2011) as well as friendship (Cheng, Caverlee, Lee, & Sui, 2011) relate to human mobility. The distinct advantage of GSMD is that it provides spatial information at up to street level precision, allowing for detailed trajectories to be extracted (Sun, Fan, Bakillah, & Zipf, 2013; Zheng, Zha, & Chua, 2012) for a wide range of applications such as point of interest classification (G. Andrienko, et al., 2013), community detection (Cranshaw, Schwartz, Hong, & Sadeh, 2012; Wakamiya, Lee, & Sumiya, 2013) and identification of anomalous movements (Chae, et al., 2015; Gabrielli, Rinzivillo, Ronzano, & Villatoro, 2014).

Research in tourism has been quick to capitalize on this feature by utilizing GSMD for estimates of inbound tourists numbers (Barchiesi, Moat, Alis, Bishop, & Preis, 2015) as well as measures of tourist activities at specific urban (Önder, Koerbitz, & Hubmann-Haidvogel, 2014) and nature based attractions (Orsi & Geneletti, 2013; Wood, Guerry, Silver, & Lacayo, 2013). GSMD also contains user information like the subject's location of origin, which is particularly useful for building character profiles (Fuchs, et al., 2013) that can be used to tailor destinations to individual interests (Jiang, Yin, Wang, & Yu, 2013). This information is also useful for destination management, especially to identify groups of destinations that compete for the same tourists (Koerbitz & Önder, 2013). Naturally, travel-planning systems also leverage on this aspect of GSMD to recommend international travel destinations (Alowibdi, Ghani, & Mokbel, 2014), nearby attractions (Zanker, Fuchs, Seebacher, Jessenitschnig, & Stromberger, 2009) and scenic routes (Chen, Shen, & Zhou, 2011; Sun, Fan, Bakillah, et al., 2013). Very few attempts have been made to characterize tourist flows with GSMD. To our knowledge, the earliest works were concerned with identifying distinctive flow patterns between popular tourist attractions (Girardin, Fiore, Ratti, & Blat, 2008) as well as to trace the type of events tourists participate in (Vaccari, et al., 2009). More recently, GSMD was utilized to compare seasonal demand for tourist accommodations (Sun, Fan, Helbich, & Zipf, 2013), investigate how the layout of cities and their tourism infrastructures influenced tourist behavior (Kadar, 2013) and reveal how destination preferences differ between demographic groups (Vu, et al., 2015).

Among the methods we have described in this section, GSMD satisfies our data quality criteria to the fullest extent (See Table 1). Similarly, previous attempts to characterize tourist flows with GSMD show good potential for this analytical approach to shed light on the RQs presented above (See Section 2.1). Nonetheless, there are some caveats associated with its use (Nabian, Offenhuber, Vanky, & Ratti, 2013; Shelton, Poorthuis, & Zook, 2015) and it is paramount to acknowledge that certain information may be omitted from the gathered data. In this respect, user penetration is the key shortcoming of GSMD as access to smartphones and the practice of geotagging is essentially limited to a narrow user demographic of technologically savvy individuals (Murdock, 2011).

Of this demographic, studies have shown that males from densely populated, urban regions are significantly overrepresented (Hecht & Stephens, 2014; Riederer, Zimmeck, Phanord, Chaintreau, & Bellovin, 2015) while ethnic minorities are marginalized (Mislove, Lehmann, Ahn, Onnela, & Rosenquist, 2011). Reliability must also be considered, as users are not obliged to provide real information (Hecht, Hong, Suh, & Chi, 2011).

4. Methodology

Geotagged Twitter data was chosen to operationalize our analysis. Apart from the distinctive advantages mentioned above, Twitter provides freely accessible mechanisms to monitor activities that occur within a specified area and timeframe. However, several technical requirements must be fulfilled in order to perform flow analysis with geotagged tweets[8]. Firstly, tweets are not directly downloadable from a web repository but must be gathered from Twitter's application programming interfaces (API) based on a well-defined set of queries. Secondly, the collected tweets in raw point format are ill structured for flow analysis and require processing prior to visualization. Finally, common flow representations are prone to visual clutter that occludes important patterns (Schulz & Schumann, 2006) and thereby require optimization for accurate representation. In this section, we provide technical details that elaborate on how data is collected, processed and visualized for communicable insights.

4.1. Data Collection

Two distinct types of data are required, namely user profiles for demographic information and geotagged tweets for spatial and temporal information. We carried out

[8] Twitter posts are colloquially referred to as tweets.

data collection in three phases. The first phase involves gathering geotagged tweets posted within the boundaries of Cilento between 29 May 2014 and 31 December 2014 through Twitter’s stream API (Twitter, 2014b). In phase two, we compiled a list of unique user Ids extracted from the tweets collected in phase one. Finally, we iteratively queried Twitter’s REST API (Twitter, 2014a) to gather geotagged tweets and user profiles that correspond to each user Id on the list compiled in phase two.

4.2. Data Processing

Our analysis requires a data structure that supports visualization and dynamic filtering of aggregated trajectories based on spatial, temporal as well as demographic constraints. For this purpose, trajectories must be extracted on a daily basis for each user Id and organized into demographic groups based on the information in each user profile. We describe our trajectory mining and tourist detection procedure below.

4.2.1. Trajectory Mining

A trajectory is a time-ordered collection of geotagged tweets that traces the movement of an individual through space and time. Each sequential pair of tweets in a trajectory is referred to as a *pathway* and reflects the movement of an individual from a location to the next. Aggregation is typically required to extract collective patterns from the data. Thus trajectories are simplified by binning the origin and destination of each pathway to cells in an $n \times m$ grid. This step reduces the large number of spatial variations among trajectories to a representative subset where the frequencies of travel along common pathways become evident.

The aggregated trajectories are expressed as a directed graph $G(V, E)$ where vertices $v_i \in V$ are cells in the grid that corresponds to physical locations in the region, while edges $e_i \in E$ indicate movement pathways between cells. We identify two types of edges. A directed edge e_{ij} is an aggregation of pathways with origins at vertex v_i and destinations at vertex v_j. A self-directed edge e_{ii} is an aggregation of pathways where both origins and destinations are vertex v_{ii}. Each edge is weighted by value f, indicating the aggregate number of trips between an origin and destination. Thus, $f\left(e_{ij}\right)$ refers to the frequency of travel between v_i and v_j.

4.2.2. Tourist Detection

To characterize the demographic features of tourist flows in Cilento, we must determine their location of origin. We obtained a demographic breakdown of the population sample by grouping individuals according to the time-zone listed on their profiles. In this manner, locals were differentiated from tourists, while foreign tourists were distinguished from domestic tourists. Finally, foreign tourists were broken down into sub groups. Time-zone was chosen instead of content from the location field because the former is formatted in a consistent manner. Twitter users choose their time-zone from a list of predefined cities, but any text information can be submitted to the location field regardless of its validity (Hecht, et al., 2011). We obtained the timezone information by systematically querying user profiles with Twitter's REST API.

Missing values were expected thus the classification was improved by clustering individuals based on the frequency of their activity in Italy. We define four metrics to accomplish this:

Td	Number of days an individual has been active in Italy.
Cd	Number of days an individual has spent in Cilento.
$\widetilde{Td}$	The median number of days that all individuals have been collectively active.
$\widetilde{Cd}$	The median number of days spent in Cilento by all individuals collectively.

The value Td indicates the total number of days an individual has created tweets, while Cd refers to the number of days that those tweets occurred within the geographic boundaries of Cilento. We consider any individual who did not declare a 'time-zone' to be domestic tourist if $Td \geq \widetilde{Td}$. Then, we define a probability index p 0.75 to distinguish domestic tourists from the locals on the basis that locals spend most of their time within the region. The probability index p is computed in the following manner:

$$p = \left\{ \begin{array}{ccc} Cd/Td & if & Cd \geq \widetilde{Cd} \\ 0 & & otherwise \end{array} \right\}$$

4.3. Data Visualization

We developed FlowSampler (Chua, Marcheggiani, Servillo, & Vande Moere, 2014), a purpose built visualization tool that enables interactive visual analysis of spatial temporal patterns in an integrated view. As shown in Figure 1, the primary interface is a flow map that depicts tourist flows among various locations in Cilento (see Figure 1a). The flow map can be dynamically filtered across four variables: Time (See Figure 1b), direction of travel, number of trips and demographic group (See Figure 1c).

Selecting a cell reveals the incoming and outgoing flows from that location (see Figure 1d).

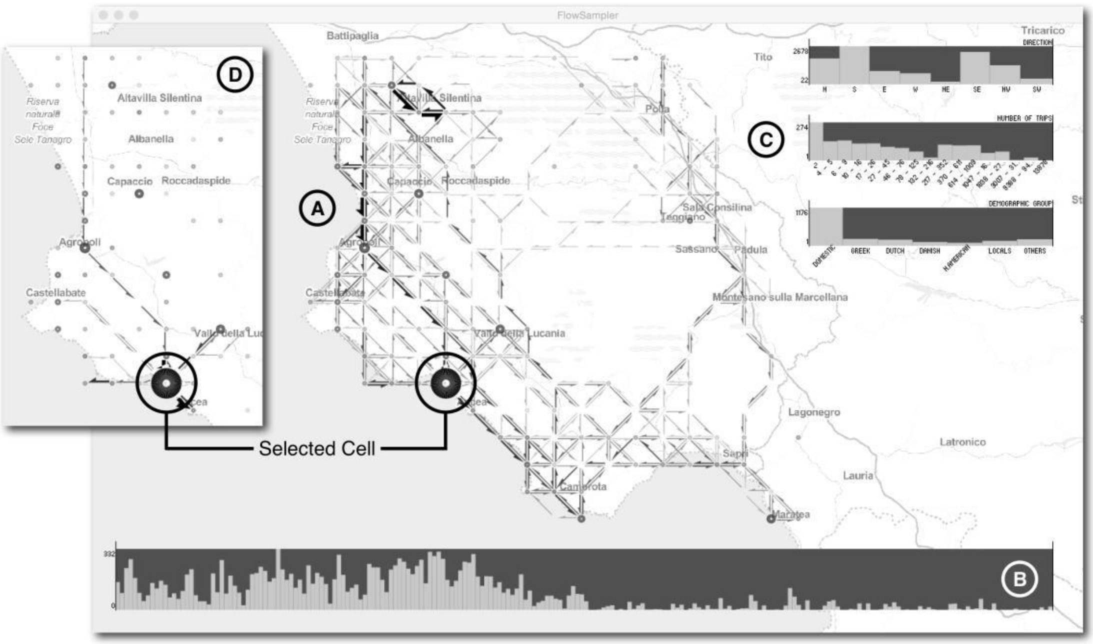
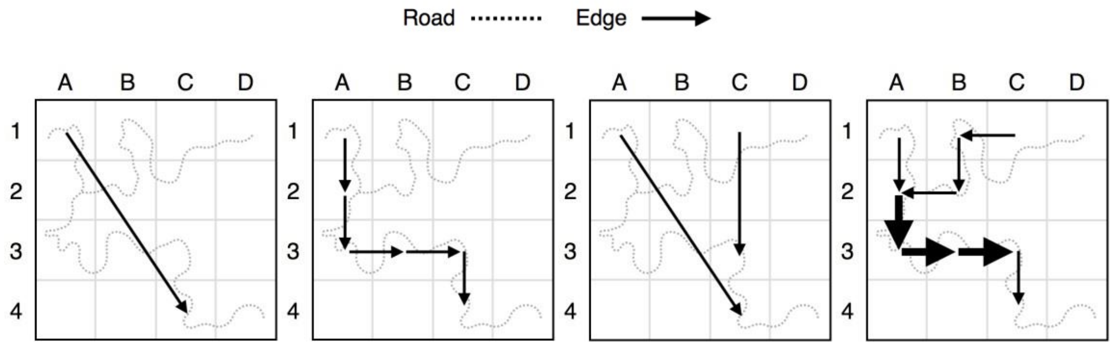


Figure 1. Components of the visualization interface. (a) Flow map geographically centered on Cilento. (b) Timeline indicating the number of unique individuals posting geotagged tweets per day. (c) Widgets for filtering the map based on direction of travel, number of trips and demographic group. (d) Cut out of the flow map depicting both incoming and outgoing flows from a selected cell.

4.3.1. Flow Map Optimization

A straightforward way to render flow maps is to draw arrows pointing from an origin directly towards a destination. The width of the arrow is often scaled to indicate a specific attribute value (e.g. Number of trips). This visual representation is easy to understand when the data is sparsely distributed, but patterns become difficult to discern when a large number of arrows intersect or overlap (Schulz & Schumann, 2006). While there are many ways to address this challenge, we employ a transition type representation (N. Andrienko & Andrienko, 2013) that divides each trajectory into segments, so that movement on a path is approximated by a sequence of transitions between adjacent discrete locations (See Figure 2). An advantage of this representation is that flows are captured as a directed graph that allows for quantitative metrics like centrality to be computed and visually compared.



(a) (b) (c) (d)

Figure 2. Visualizing edges on the flow map. (a) Straight forward representation of an edge where an arrow is drawn between the origin and destination. (b) Shortest route representation of the edge shown in (a) based on the reference road network. (c) Straight forward representation of two edges. (d) Shortest route representation of two edges as illustrated in (c) where the thicker arrows depicts movement along a common path.

4.3.2. Route Identification

To identify possible travel routes, we render the shortest path between a pair of cells based on a reference road network. This is preferred over more sophisticated routing techniques, as it is straightforward to explain and simple for a lay audience to understand. A diagrammatic explanation is shown in Figure 2. An arrow is drawn between vertex A1 and C4 to represent the edge in a straightforward way (see Figure

2a). The same edge is represented in Figure 2b by a sequence of arrows that depict the shortest route between both cells. In this instance, the widths of the arrows are scaled according to the aggregate amount of movement along a common path. Figure 2c and 2d illustrate a scenario whereby the shortest routes between two pairs of cells converge at cells A3, B3 and C3.

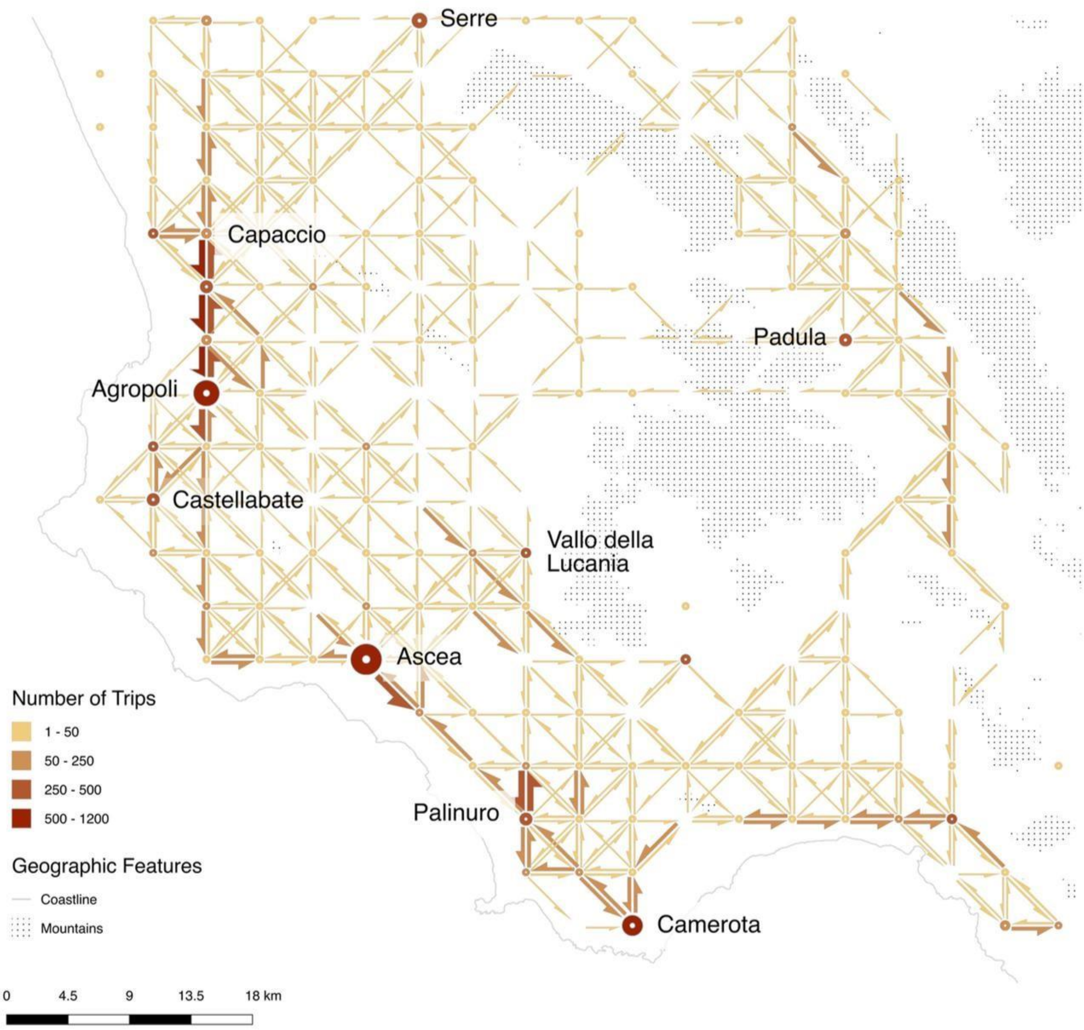


Figure 3. Visualization of flows derived from 3,135 individuals who have posted at least one tweet in Cilento.

5. Results

Our data consisted of 72,031geotagged tweets posted by 3,135 unique individuals. On average, there were 193 (SD = 2.639) tweets per trajectory. The low standard deviation indicates that majority of the trajectories contain a relatively similar number of tweets. Figure 3 provides a summary of the flows in Cilento binned to a 19 × 20 grid. Each grid cell measured 4.5 × 4.5 kilometers. This resolution was chosen to closely map every settlement to a distinct cell. The width and color of the edges were binned to a scale consisting of four value ranges. The map depicts more activity along the coastline and reveals several important hubs in the region. There are several high frequency pathways that indicate important flows around Ascea and Palinuro but the series of movements between Capaccio and Agropoli is most distinctive.

5.1. Demographic Breakdown of Tourist

We detected 138 (4.4%) locals, 1176 (37.5%) domestic tourists and 628 (20%) foreign tourists. 1193 (38.1%) individuals did not meet the classification conditions defined above and excluded from further analysis. Figure 4 depicts the demographic breakdown according to location of origin. Locations outside of Europe were aggregated into wider geographic regions to simplify the classification. Our classification indicates that Cilento primarily attracts domestic tourist. Correspondingly, Greek (5.4%) and Dutch (5%) tourist account for more than half of foreign tourism. The discovery of tourists originating from other locations suggests that the region attracts a diverse audience.

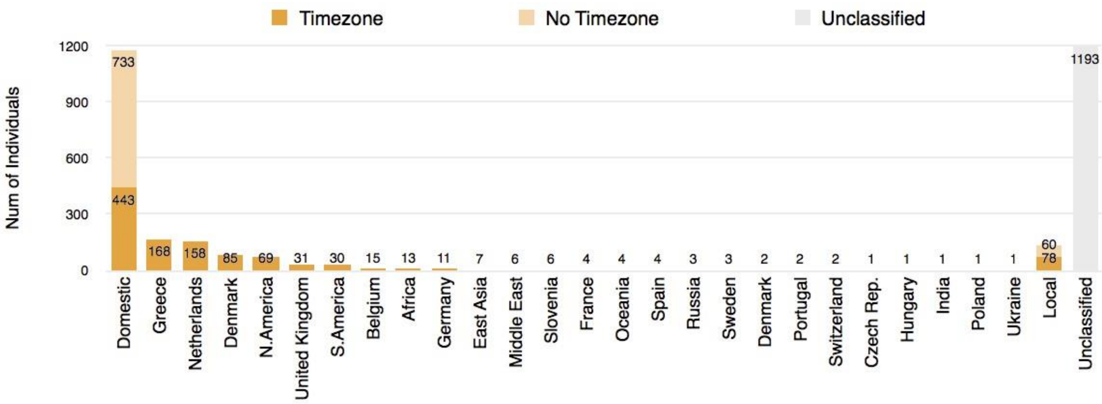


Figure 4. Demographic Breakdown of individuals in Cilento. Locations outside of Europe have been aggregated into wider geographic regions.

5.2. Uncovering Temporal Characteristics of Tourism

We compared variations in Twitter usage among demographic groups to uncover the temporal features of tourism in the region. Figure 5 indicates the aggregate number of people who posted geotagged tweets on a daily basis. The number of locals is depicted in grey, while tourist numbers are represented with three colored time series: An aggregated time series for tourists in general, and two other time series for to facilitate comparison between domestic and foreign tourists. The data shows an increase in tourist activity in late June followed by a peak in the second week of August. The trend declines by the first week of September. Incidentally, we observe a reverse trend among the locals during that period. The distinctive difference between domestic and foreign tourism is the timeframe during which they occur. This temporal pattern is characterized by the decline in foreign tourists numbers just as the presence of domestic tourist begin to escalate and peak. The bimodal distribution observed in the aggregate time series is also explained by this pattern. The highest number of tweets posted by foreign tourist is recorded during first week of July. A similar observation for domestic tourists occurs at the first week of August.

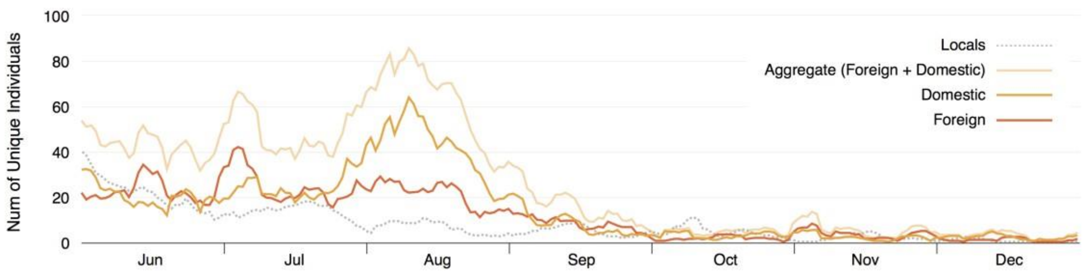
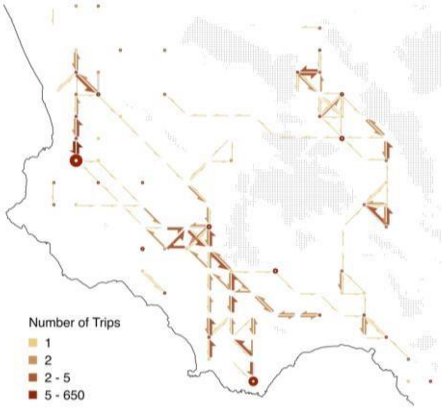


Figure 5. Temporal analysis of local and tourists twitter usage activity.

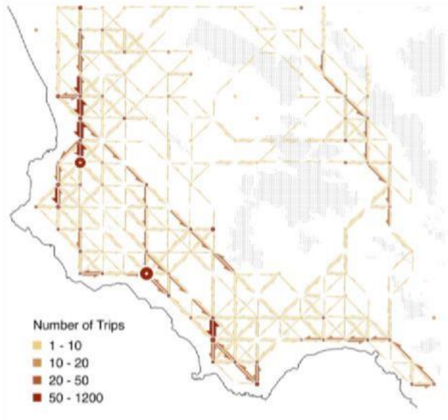
5.3. Spatial Topology of Tourist Flows

5.3.1. Circulation

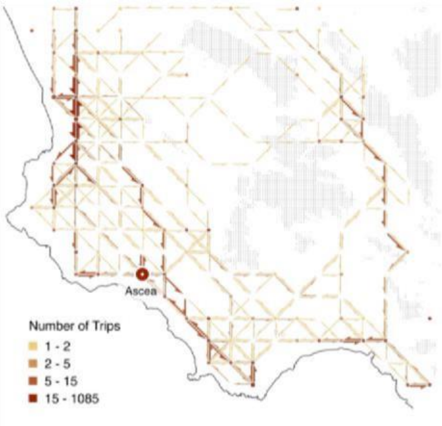
Circulation refers to the diffusion of flows in a system. Figure 6 presents a visual comparison of how individuals from various demographic groups move throughout the region. Figure 6a depicts the movements of locals while Figure 6b illustrates the aggregated tourist flows. Comparison between both maps reveals a substantial difference in the way individuals from both demographic groups circulate the region. In contrast to locals who primarily move inland, tourist activity tends to be situated along the coastline. The route between Capaccio and Agropoli however, appears to be equally important for both groups. Further comparisons between domestic (See Figure 6c) and foreign (See Figure 6d) tourists reveal several spatial differences. Whereas foreign tourists tend to be situated around Agropoli and Palinuro, domestic tourists are seen in Ascea. The routes taken by domestic tourists also differ substantially from foreign tourists in that they are situated further inland. Turning our attention solely on foreign tourism, we discover that Greek (See Figure 6e) and Dutch tourist, (See Figure 6f) travel to a diverse set of locations while those from Denmark (See Figure 6g) and North America (See Figure 6h) are limited to a smaller number of destinations.



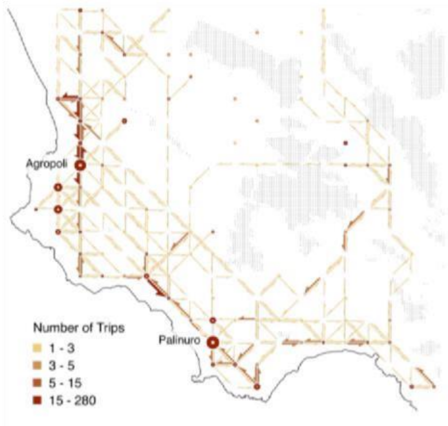
(a) Locals



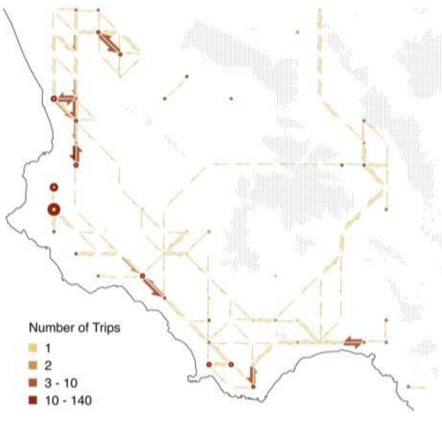
(b) All Tourist Aggregated



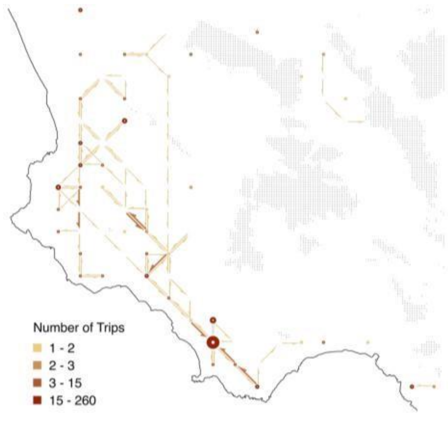
(c) Domestic Tourists



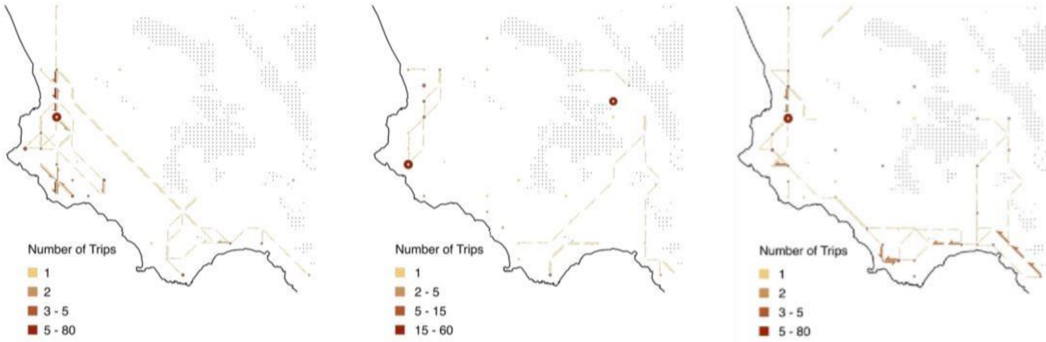
(d) All Foreign Tourists Aggregated



(e) Greek Tourists



(f) Dutch Tourist

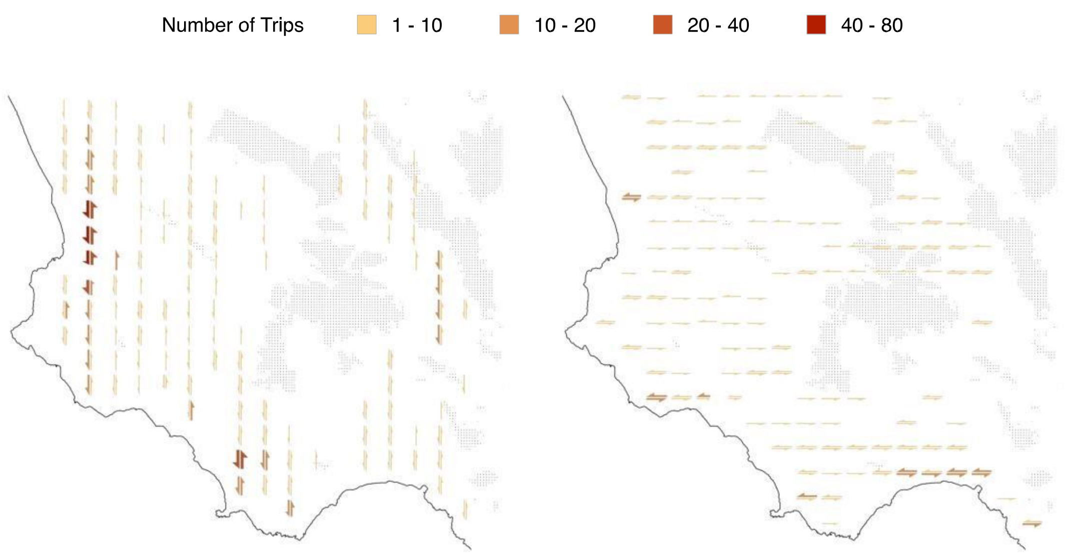

Number of Trips
1
2
3 - 5
5 - 80
Number of Trips
1
2 - 5
5 - 15
15 - 60
Number of Trips
1
2
3 - 5
5 - 80

(g) Danish Tourists (h) N. American Tourists (i) Other Foreign Tourists

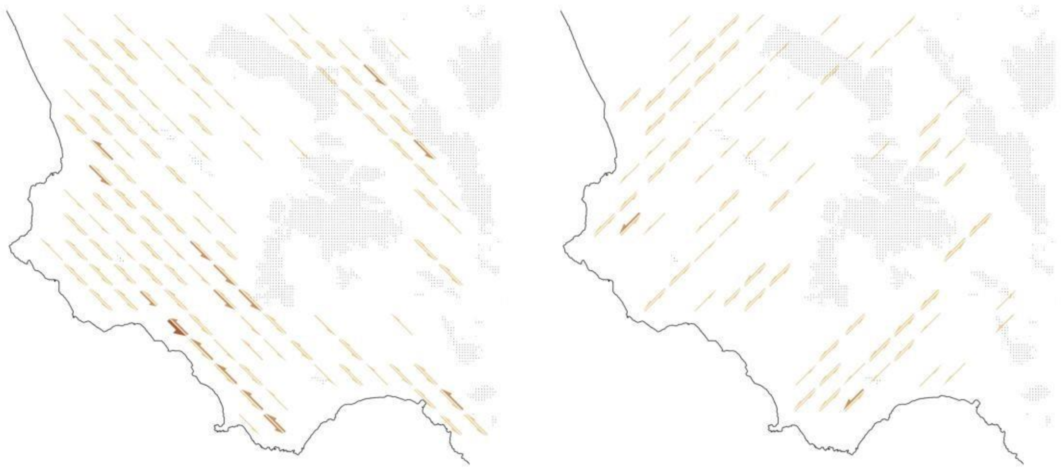
Figure 6. Comparing movements of tourists from different countries of origins.

5.3.2. Direction

Directionality is another spatial feature of movement. In Figure 7, we show the aggregated tourist flows divided into four directions to compare the frequency of travel per direction: N to S (See Figure 7a), E to W (See Figure 7b), NW to SE (See Figure 7c) and NE to SW (See Figure 7d). Self-directed flows have been removed from the map, as they do not provide directional information. The maps reveal that tourists primarily travel in a S-SE direction along the coastal settlements. To a lesser extent, traces of N-NW bound travel in the opposite direction can be detected along the same route. Filtering the map along the timeline indicates that northbound flows are traces of egress that occur at the end of August when tourist depart.



(a) North-South (b) East-West



(c) North West-South East (d) North East-South West

Figure 7. Comparing the directionality of all aggregated tourist flows along four directions.

5.3.3. Centrality

The centrality of tourist attractions can be determined by its popularity among tourists and is assessed in two stages. First, popular attractions were identified based on the unique number of tourists recorded in each cell of the grid. Thereafter, the influence of these attractions is evaluated in comparison with less popular venues on the basis of connectivity; that is the geographic spread of inward and outward tourist flows to other location in the region. Each cell in Figure 8 is colored in a shade of orange to indicate its distance from the transport infrastructure and tinted in a shade of grey to represent the unique number of tourists who posted tweets within its boundaries. Popular attractions that recorded more than 90 (Q_3) unique visitors appear in the darkest shade of brown – a blend of orange and grey. Unsurprisingly, these attractions are located along the coastline where public transport is readily available. Additionally, we observe that the number of tourists in a given cell decreases as its distance from the road or rail network increases. Comparison of the connectivity between popular attractions to other venues indicates an asymmetry in spread (See Figure 9). Whereas popular attractions appear well connected to other locations (See Figure 9a), movement inland is limited to adjacent localities (See Figure 9b). Tracing movements over time reveals that tourists constantly travel in an "inland, coastal, inland" sequence in order to move between two disjoint inland locations.

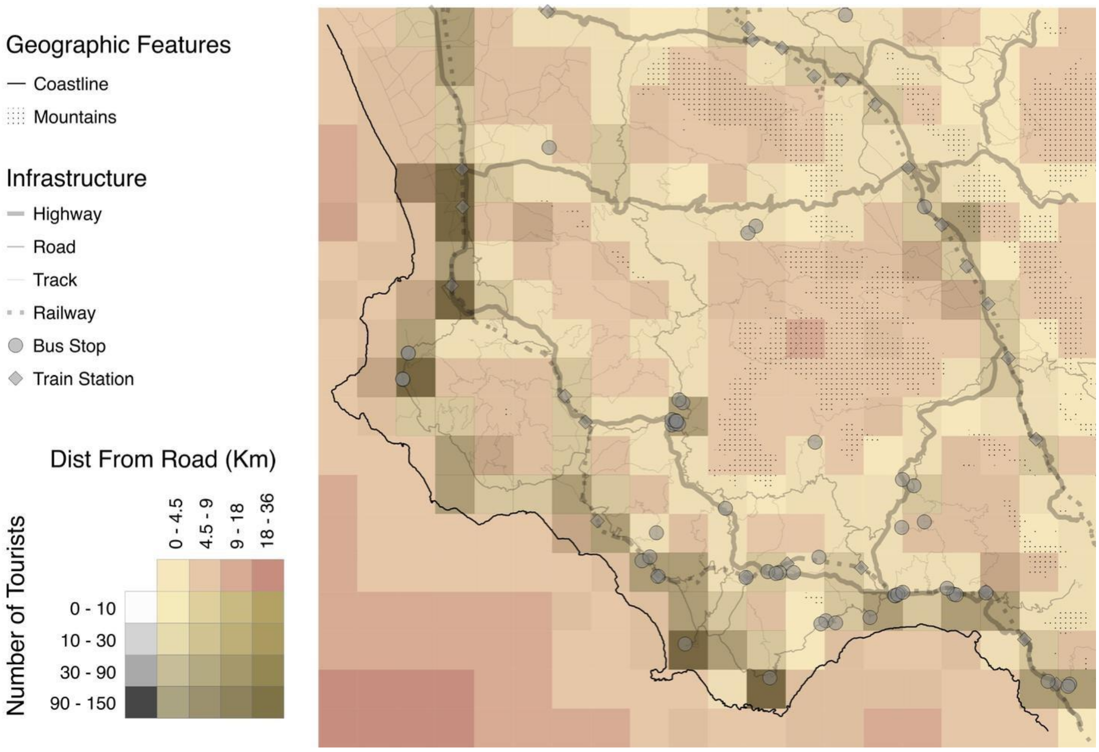
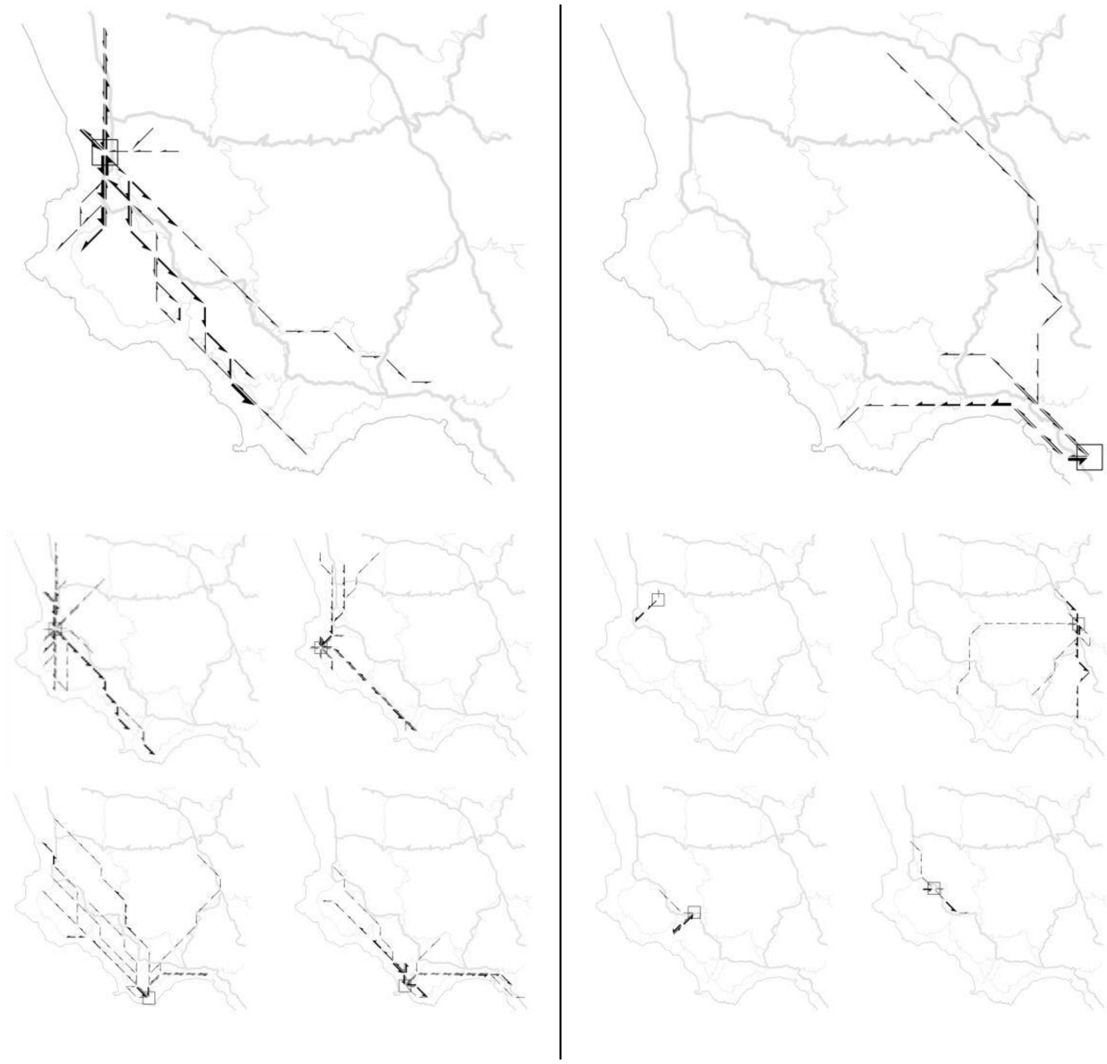


Figure 8. Popular locations in Cilento based on the total number of unique visitors in each cell of the grid.



(a) Most Popular Locations (b) Other Locations

Figure 9. Popular locations defined on the basis of connectivity. (a) Locations along the coastland are relatively well connected. (b) Inland flows are limited to adjacent localities.

5.4. Insights

In this sub-section, we review our findings structured around the initial RQs and describe our interpretation of tourist flow patterns in the region based on the information obtained.

5.4.1. (RQ1) What are the meaningful tourist profiles in the region?

There is a greater presence of domestic rather than foreign tourists in Cilento (x^2=166.4656, p<0.01). According to our results, this finding is in agreement with figures from our current understanding of tourist Flows (see Table 3). The breakdown of foreign visitors into smaller groups indicates diverse tourist demographics. In particular, the Greek and Dutch tourist account for more than half of foreign tourism (x^2=17.1221, p<0.01). Danish and North American tourist also have a sizable presence while remaining visitors originate from other locations in Europe, East Asia, South America the Middle East and Africa.

5.4.2. (RQ2) What are the valuable patterns of tourist flows in the region?

We have observed several dominant spatial-temporal patterns in tourist flows. Temporal analysis of Twitter usage reveals that foreign and domestic tourism does not occur concurrently but in overlapping stages that peak in different moments in time. Similarly, there is a substantial difference in the way locals and tourists circulate the region (t(352)=28.892, p<0.01). Tourists generally travel along the coastline while locals travel further inland. Tourists generally travel from the northwestern region of Cilento heading in a southerly direction along the coastline towards destinations in the south. This pattern of movement is prevalent among foreign and domestic tourists but is not observed among locals who, on the contrary, travel northwards. Further comparison between different groups of foreign tourist indicates two distinct patterns of movement. Greek and Dutch tourists are observed to be more mobile than their counterparts in that they travel longer distances while tourists from Denmark and North America are limited to a smaller number of destinations. Finally, the recurrence of “inland, coastal, inland” travel is particularly striking in that it reflects the importance of the coastal settlements as transit hubs in the overall transport infrastructure.

5.4.3. (RQ3) Where are the tourist attractions in the region and how do they differ?

We identified six popular attractions that spatially correspond to settlements in the region (See Figure 9a). These attractions are located along the coastline where public transport is readily available. Access to public transport appears to be paramount factor that affects the popularity of attractions in Cilento since Vallo Della Lucania and Padula, two UNESCO heritage sites situated inland, did not receive as many visitors as Paestum, their counterpart along the coast. This pattern is prevalent among foreign tourist and less pronounced with domestic tourists who were likely to have driven into the region by car in on the highway situated inland. Comparatively, foreign tourists travel along the coastline where the bus and train service is directly accessible.

5.4.4. Interpretation

Tourist flows in Cilento is likely to be defined by the physical configuration of the existing transport infrastructure. Because public transportation does not extend beyond the coast, tourists are required to obtain personal transport for inland travel. This insight provides evidence for the relative popularity of the coastline and the lack of tourist movement inland. The southward pattern of flow is likely to result from this configuration as points of entry into Cilento, from regional transport hubs like Serre and Capaccio are located at the northern half of the region while a scenic beach at Camerota and a heritage site at Vallo di Diano are located at the southern half of the region. Based on these findings, we believe that new economic opportunities can be created by expanding the transport infrastructure inland. A permanent transport system is costly to run. However, tailoring the service to a specific tourist demographic within the duration of seasonal peaks, may lower operating cost and mitigate this issue. Since foreign tourists are regarded to have greater financial spending power but are spatially bounded to the coastline, a seasonal transportation service primarily catering to foreign tourists is a feasible policy solution. Furthermore, tailored services like tours to inland attractions can be organized and marketed to match the destination preference of different tourist demographic groups. To this end, formal studies are required to elaborate on how the existing transport infrastructure affects tourism.

5.5. Discussion

Existing knowledge of tourist flows in Cilento is derived from the analysis of regional and provincial data, direct observation in the field or extracted from official reports, publications and anecdotes from local staff and stakeholders. While sufficient to broadly understand the context and challenges within the region, such information does not meet our data quality criteria as outlined in Table 2. In this regard, our analytical approach based on geotagged tweets demonstrates that GSMD is a valuable source of tourist flow information. As shown in Table 3, the insights we obtained extend the current understanding of tourist flow patterns. Specifically, our approach increased the detail in spatial, temporal and demographic information available. In comparison to the repertoire of methods that are currently in place, our approach is able to monitor tourist movements across large geographic areas without the need for dedicated physical sensing infrastructures. Though there were previous attempts to characterize tourist flows with GSMD, these were limited to individual features of flow and generally situated within cities. Conversely our analytical approach provides relatively sophisticated descriptions of movement as well as profiles of tourists over large a region. In this respect, our analytical approach presents a substantial advancement in describing tourist flows with GSMD.

Nonetheless, there are limitations to be aware of when drawing conclusions from insights derived from analyzing GSMD since our approach is principally driven by data provided at the social media users' discretion, without ground truth for verification. The results might be somewhat misleading if biases in the data are unaccounted for. As described earlier, demographic studies have shown that a large part of Twitter users are young adults, and potentially represent only a partial slice of the actual tourist population on the ground. Moreover, the data may not capture complete travel itineraries, as mobile devices are less likely to used when engaging in certain types of activities. It is also it is unclear if cellular or GPS signal strength affect social media usage or geotagging respectively. From this point of view, it is crucial to acknowledge that the insights may overemphasize particular tourist demographics, activities and attractions. We also expect the use of GSMD to raise privacy and ethical concerns related to the collection of data without direct consent from social media users. On this subject, it should be expressed that the data gathered includes only information that users explicitly disclose. Furthermore, the data is aggregated in a way that all traces of individual trajectories are removed.

Our work constitutes an integrated approach for tourist flow analysis with limited consideration of related computational methods. Trajectories are mined as observed movements between origin and destinations without the actual travel paths in between. We introduced route identification to address this aspect of movement yet this information is purely inferred based the on shortest path between two locations that might be very different from the actual path taken. In this respect, more sophisticated routing techniques based on criteria such as well known points of interest could be implemented to obtain higher information accuracy. To make large analysis computationally tractable, we aggregated the data by dividing the observation area into grid of cells. As a consequence, the finest unit of analysis occurs on the level of cells. In this respect, movements within cells, although represented in our visualization, were not further investigated. Naturally, these cells contain urban areas within its boundaries and can be further subdivided for finer breakdowns of flow patterns. Alternatively, quadtree or kd-tree partitioning can be applied to derive grids that distinguish between urban and rural areas, based on a supplemental dataset like the whereabouts of buildings to better reflect urban density. Accordingly, our approach to demographic classification can also be improved. Considering that a substantial number of Twitter users could not be classified on the basis of time-zone, localization algorithms can be employed to determine their location of origin on the basis of content from the "location" field on their individual profiles.

Despite these limitations, we would like to point out that our approach provides equally valuable and alternative insights that are complimentary to the current understanding of tourist flows in Cilento, derived from existing data sources. Additionally, we do not claim that it is an improvement on, or replacement for, other approaches to characterize tourist flows. While our findings are specific to Cilento, much of what has been found may be transferred to other regional destinations though the local context should always be taken into account. In this regard, future attempts at similar analysis may consider expanding the source data beyond geotagged tweets to include other GSMD like geotagged photos, to capture a wider spectrum of the tourist demographics and activities. Likewise, official tourism statistics maybe incorporated into the analytical procedure to uncover potential biases in the data. Our work has focused on the spatial, temporal and demographic features of tourist flows. Yet the embedded text content in each tweet has been left out. In this regard, semantic and sentiment analysis may offer new perspectives and provide rich contextual information for certain activities that tourist engage in. Since GSMD is generally accessible through public APIs, large datasets can be obtained at relatively low costs. Thus, similar analysis can be undertaken by smaller organizations that have limited resources at their disposal when tools for data collection, processing and visualization exist. Correspondingly, our analytical approach can be useful for public administration and large tourism enterprises to develop indicators for sustainable planning of territorial resources and benchmarking markets respectively.

Table 3. Comparison of information from the analysis of GSMD to the current understanding of tourist flows.

Findings Based on Geotagged Social Media Data			Current Understanding of Tourist Flows
Features		Insights	
Demographic		x Largely patronized by domestic tourists (67%). x Foreign tourists (33%) originate from many locations but Greek (9%), Dutch (8%), North American (4%) and Danish (4%) are most prominent. The remaining (8%) originate from other locations in Europe, East Asia, South America the Middle East and Africa.	x More domestic (63%) than foreign (37%) tourists. x Anecdotes suggest that majority of the tourist are Dutch and Greek.
Spatial	Circulation	x Limited mobility inland. x Tourists transit along the coastline to travel long distances.	No Information

	Directionality	x Tourists primarily travel in a southerly direction passing through the coastal settlements on the road or rail network. x Southward flow due to the configuration of transport where transit hubs are located in the north of the region while the tourist attractions are located in the south.	x No formal studies of tourist movements till date. x Anecdotal evidence regarding mode of transport suggest that domestic tourists drive to their destinations while foreign tourist journey to Cilento by train, where they alight at either Capaccio or Vallo di Lucania. Subsequent trips towards various locations in the region are then made by bus.
	Centrality	Popular tourist attractions are located along the coast and have immediate access to the transport infrastructure. As a result, these locations are better connected than those situated inland.	x Currently no consensus on any form of ranking. x Anecdotes suggest that individual municipalities claim to be more important than others.
Temporal		Analysis reveals a bimodal trend where foreign and domestic tourist activities occur over different durations and peak at separate moments in time.	Official tourists season begins on the 2nd week of May till the end of August.

6. Conclusion

We have described a set of findings from studying tourist flows through the lens of GSMD. Our approach - developing an analytical technique to collect and investigate the spatial, temporal and demographic features of tourist flows, enables relatively sophisticated descriptions of tourist movement, as well as the demographic profiles of tourist groups. However, biases in the data as well as methodological limitations should be considered when drawing conclusions from analysis of GSMD.

Nonetheless, this is the first large-scale observational study of tourist flows that to our knowledge attempt to provide a comprehensive description of tourist profiles and their associated movement.

7. References

Alowibdi, J. S., Ghani, S., & Mokbel, M. F. (2014). VacationFinder: a tool for collecting, analyzing, and visualizing geotagged Twitter data to find top vacation spots. In *Proceedings of the 7th ACM SIGSPATIAL International*

Workshop on Location-Based Social Networks, 9-12.

Andrienko, G., Andrienko, N., Fuchs, G., Raimond, A.-M. O., Symanzik, J., & Ziemlicki, C. (2013). Extracting semantics of individual places from movement data by analyzing temporal patterns of visits. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Computational Models of Place*, 9-15, Orlando, FL, USA.

Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization, 12*(1), 3-24.

Ashley, C., De Brine, P., Lehr, A., & Wilde, H. (2007). *The role of the tourism sector in expanding economic opportunity*: John F. Kennedy School of Government, Harvard University.

Bakillah, M., Lauer, J., Liang, S., Zipf, A., Jokar Arsanjani, J., Loos, L., & Mobasheri, A. (2014). Exploiting big VGI to improve routing and navigation services. In H. A. Karimi (Ed.), *Big Data Techniques and Technologies in Geoinformatics* (pp. 177-192). Boca Raton, FL, USA: CRC Press, Taylor & Francis Group.

Barchiesi, D., Moat, H. S., Alis, C., Bishop, S., & Preis, T. (2015). Quantifying international travel flows using Flickr. *PloS one, 10*(7), 1-8.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics, 214*(1), 481-518.

Beecham, R., Wood, J., & Bowerman, A. (2014). Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems, 47*, 5-15.

Chae, J., Cui, Y., Jang, Y., Wang, G., Malik, A., & Ebert, D. S. (2015). Trajectorybased visual analytics for anomalous human movement analysis using social media. In *Proceedings of the EuroVis Workshop on Visual Analytics*, 43-47, Cagliari, Sardinia, Italy.

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. In *Proceedings of the 27th International IEEE Conference on Data Engineering (ICDE) 2011* 900-911.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 81-88, Barcelona, Spain.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1082-1090, San Diego, CA, USA.

Christaller, W. (1964). Some considerations of tourism location in Europe: the peripheral regions - underdeveloped countries - recreation areas. *Papers in Regional Science, 12*(1), 95-105.

Chua, A., Marcheggiani, E., Servillo, L., & Vande Moere, A. (2014). FlowSampler: visual analysis of urban flows in geolocated social media data. In L. M. Aiello & D. McFarland (Eds.), *Social Informatics* (pp. 5-17) Springer.

Cooper, C. (1981). Spatial and temporal patterns of tourist behaviour. *Regional Studies*, *15*(5), 359-371.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. M. (2012). The livehoods project: utilizing social media to understand the dynamics of a city. In

Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 58-65, Trinity College, Ireland.

Ferreira, N., Poco, J., Vo, H. T., Freire, J., & Silva, C. T. (2013). Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2149-2158.

Fuchs, G., Andrienko, G., Andrienko, N., & Jankowski, P. (2013). Extracting personal behavioral patterns from geo-referenced tweets. In *Proceedings of the 16th AGILE Conference on Geographic Information Science*, Leuven, Belgium.

Gabrielli, L., Rinzivillo, S., Ronzano, F., & Villatoro, D. (2014). From tweets to semantic trajectories: mining anomalous urban mobility patterns. In J. Nin & D. Villatoro (Eds.), *Citizen in Sensor Networks* (pp. 26-35) Springer.

Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Digital footprinting: uncovering tourists with user-generated content. *IEEE Pervasive Computing*, *7*(4), 36-43.

Girardin, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services*, *2*(1), 41-56.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779-782.

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.

Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart:

the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 237-246, Vancouver, Canada.

Hecht, B., & Stephens, M. (2014). A tale of cities: urban biases in volunteered geographic information. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, 197-205, Ann Arbor, Michigan, USA.

Jiang, K., Yin, H., Wang, P., & Yu, N. (2013). Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing, 119*, 17-25.

Kadar, B. (2013). Differences in the spatial patterns of urban tourism in Vienna and Prague. *Urbani izziv, 24*(2), 96-111.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1-14.

Koerbitz, W., & Önder, I. (2013). Destination benchmarking with geotagged photographs. In Z. Xiang & I. Tussyadiah (Eds.), *Information and Communication Technologies in Tourism 2014* (pp. 201-211) Springer.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: the geography of Twitter. *First Monday, 18*(5).

Lew, A., & McKercher, B. (2006). Modeling tourist movements: a local destination analysis. *Annals of Tourism Research, 33*(2), 403-423.

Mansfeld, Y. (1990). Spatial patterns of international tourist flows: towards a theoretical framework. *Progress in Human Geography, 14*(3), 372-390.

Mavoa, S., Oliver, M., Witten, K., & Badland, H. M. (2011). Linking GPS and travel diary data using sequence alignment in a study of children's independent mobility. *International Journal of Health Geographics, 10*(1), 1-10.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the demographics of Twitter users. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 554-557, Barcelona, Spain.

Murdock, V. (2011). Your mileage may vary: on the limits of social media. *SIGSPATIAL Special*, *3*(2), 62-66.

Nabian, N., Offenhuber, D., Vanky, A., & Ratti, C. (2013). Data dimension: accessing urban data and making it accessible. *Proceedings of the Institution of Civil Engineers - Urban Design and Planning, 166*(1), 60-75.

Nagatani, T. (2002). The physics of traffic jams. *Reports on Progress in Physics, 65*(9), 1331.

Neuhaus, F. (2010). UrbanDiary - A tracking project capturing the beat and rhythm of the city: using GPS devices to visualise individual and collective routines within Central London. *The Journal of Space Syntax, 1*(2), 336.

Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2014). Tracing tourists by their digital footprints. The case of Austria. *Journal of Travel Research*, 1-8.

Orsi, F., & Geneletti, D. (2013). Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *Journal for Nature Conservation, 21*(5), 359-368.

Prideaux, B. (2000). The role of the transport system in destination development. *Tourism management*, *21*(1), 53-63.

Riederer, C., Zimmeck, S., Phanord, C., Chaintreau, A., & Bellovin, S. M. (2015). “I don’t have a photograph, but you can have my footprints." – Revealing the demographics of location data. In *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media*, University of Oxford, Oxford, UK.

Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011). Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS one, 6*(1).

Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: evidence from a six-week travel diary. *Transportation, 30*(1), 13-36.

Schulz, H. J., & Schumann, H. (2006). Visualizing graphs - A generalized view. In *Proceedings of the 10th International Conference on Information Visualization*, 166-173, London, England.

Sevtsuk, A., & Ratti, C. (2009). Does urban mobility have a daily routine? Explorations using aggregate mobile network data. In *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*, Hong Kong, China.

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning, 142*, 198-211.

Shoval, N., & Isaacson, M. (2009). Methodological aspects of measurement and visualization of tourists’ spatial behavior. In S. Page (Ed.), *Tourist Mobility and Advanced Tracking Technologies* (Vol. 19, pp. 28-43) Routledge.

Sieber, R. (2006). Public participation geographic information systems: a literature review and framework. *Annals of the Association of American Geographers, 96*(3), 491-507.

Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018-1021.

Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2013). Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems*.

Sun, Y., Fan, H., Helbich, M., & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: using Flickr to analyze spatial and temporal pattern of tourist accommodation. In J. M. Krisp (Ed.), *Progress in Location-Based Services* (pp. 57-69) Springer.

Sykora, L., & Mulicek, O. (2014). Functional analysis of urban systems: identification of small and medium sized towns and their territorial arrangements. In L. Servillo (Ed.), *TOWN. Small and medium sized towns in their functional territorial context. Scientific Report* (pp. 113-161). Luxembourg: ESPON.

Thornton, P., Williams, A., & Shaw, G. (1997). Revisiting time-space diaries: an exploratory case study of tourist behaviour in Cornwall, England. *Environment and Planning A, 29*(10), 1847-1867.

Twitter. (2014a). REST APIs. retrieved Novemember 2014, from http://dev.twitter.com/rest/public.

Twitter. (2014b). The streaming APIs. retrieved November 2014, from http://dev.twitter.com/streaming/overview.

Vaccari, A., Liu, L., Biderman, A., Ratti, C., Pereira, F., Oliveirinha, J., & Gerber, A. (2009). A holistic framework for the study of urban traces and the profiling of urban processes and dynamics. In *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems.*, 1-6.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism management, 46*, 222-232.

Wakamiya, S., Lee, R., & Sumiya, K. (2013). Social-urban neighborhood search based on crowd footprints network. In A. Jatowt, E. P. Lim, Y. Ding, A. Miura, T. Tetzuka, G. Dias, K. Tanaka, A. Flanagin & B. T. Dai (Eds.), *Social Informatics* (pp. 429-442) Springer.

Williams, S. (1998). Tourism, geography and geographies of tourism. In *Tourism Geography* (2 ed., pp. 3-25) Psychology Press.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports, 3*(2976), 1-7.

WTTC. (2015). Travel & tourism economic impact 2015. World. retrieved November 2015, from http://www.wttc.org/research/economic-research/economic-impact-analysis/.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism management, 31*(2), 179-188.

Zanker, M., Fuchs, M., Seebacher, A., Jessenitschnig, M., & Stromberger, M. (2009). An automated approach for deriving semantic annotations of tourism products based on geospatial information. In *Proceedings of the International Conference on Information and Communication Technologies in Tourism*, 211-221, Amsterdam, Netherlands.

Zheng, Y.-T., Zha, Z.-J., & Chua, T.-S. (2012). Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology, 3*(3).

Appendix 1. Survey of the existing practice in each municipal of Cilento.

Municipals of Cilento	What are the existing methods used to gauge tourist demographic?	What are the existing methods used to determine tourist movements?	What are the existing methods used to detect tourist hotspots?	How is the popularity of tourist attractions benchmarked?
Agropoli	ISTAT* or on site interviews at public activities and receptions.	ISTAT* and surveys at tourism offices.	Information on the internet.	Opinion polls
Alfano	None	None	ISTAT* or EPT[§]	None
Ascea/Velia	EPT[§]	EPT[§]	Chamber of Commerce Periodical	None
Camerota	ISTA and EPT[§]	ISTAT*, EPT[§] and the Campania Region periodical.	Sectorial publications	Registration at the entrance of cultural and natural attractions.
Campora	None	Surveys at tourism offices when time permits.	None	None
Cannalonga	None		None	None
Casaletto Spartano	Ticket sales at local attractions.	Ticket sales at local attractions.	None	Ticket sales at local attractions.
Castelcivita	Observation among the caves situated within our municipality.	ISTAT*	ISTAT*	Ticket sales or registration at local attractions.
Castellabate	ISTAT*, surveys at tourism offices and check-ins to travel accommodations.	EPT[§]	EPT[§]	Ticket sales at the castle of the Abate and the museum of Sacra.
Centola/ Palinuro	Check-ins at travel accommodations.	On site interviews at public activities and receptions.	None	Surveys at various travel accommodations.
Ceraso	None	None	Observation	None
Cuccaro Vetere	None	ISTAT*	None	None
Gioi	None	None	None	None
Ispani	ISTAT* and check-ins to travel accommodations.	EPT[§]	None	None
Laurito	None	None	None	None
Moio della Civitella	None	Return tourism through database of second homes.	None	None
Montano Antilia	None	None	None	None.
Montesano sulla Marcellana	Ticket sales at two museums in the city of Montesano.	None	Interviews at two museums in the city of Montesano	None
Morigerati	WWF Periodicals	WWF Periodicals	None	None
Pertosa	Observation	Check-ins at travel accommodations.	None	None
Piaggine	None	None	Formal discussions organized by the Cilento Park Authority.	None
Rofrano	Check-ins at travel accommodations.	None	None	None
Roscigno	None	None	Surveys at tourism offices.	None
Rutino	None	None	None	None
San Rufo	ISTAT*, Chamber of Commerce Periodical and Reports from Cilento Park Authority.	None	Surveys at tourism offices, on site interviews at public activities and receptions.	Surveys at tourism offices, on site interviews at public activities and receptions.
Santo Angelo a Fasanella	Tour registrations at the caves of St. Angelo a Fasanella	Observations and on site interviews.	None	None
Sapri	ISTAT* and EPT[§]	Formal discussions with local stakeholders.	None	Formal discussions with local stakeholders.
Serramezzana	EPT[§], observations and checkins at travel accommodations.	EPT[§] and reports by tour operators.	EPT[§]	None
Teggiano	Ticket sales at the Diocesan museum.	Ticket sales at the Diocesan museum.	None	Ticket sales at the Diocesan museum.
Torraca	None	None	None	None

Tortorella	EPT[§], check-ins at travel accommodations.	EPT[§]	None	Ticket sales at local attractions.

* Istituto Nazionale di Statistica, Italian National Institute of Statistics

§ Ente Provinciale per il Turismo, Sarleno Provincial Agency for Tourism

Figure 1

Click here to download high resolution image

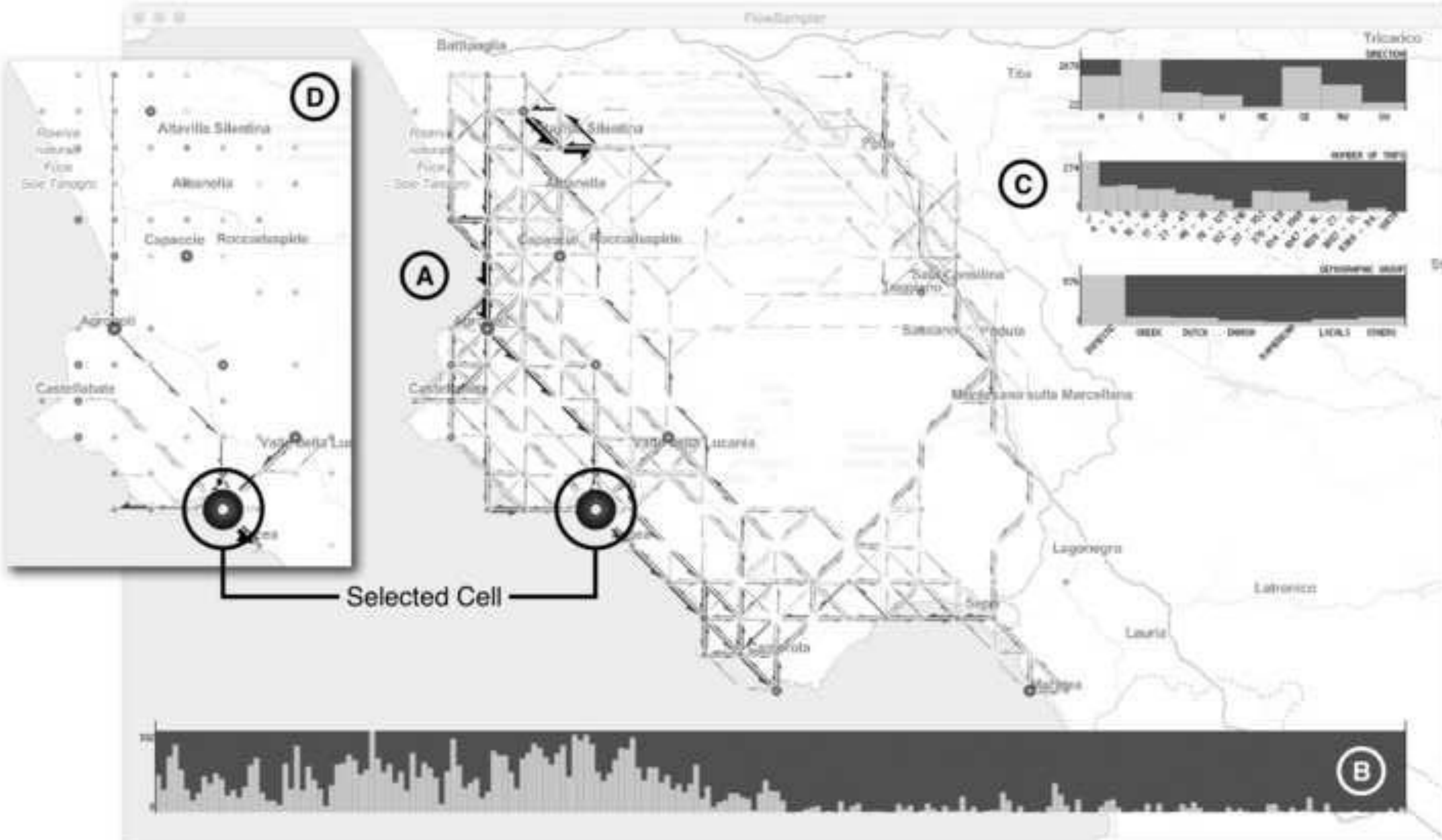
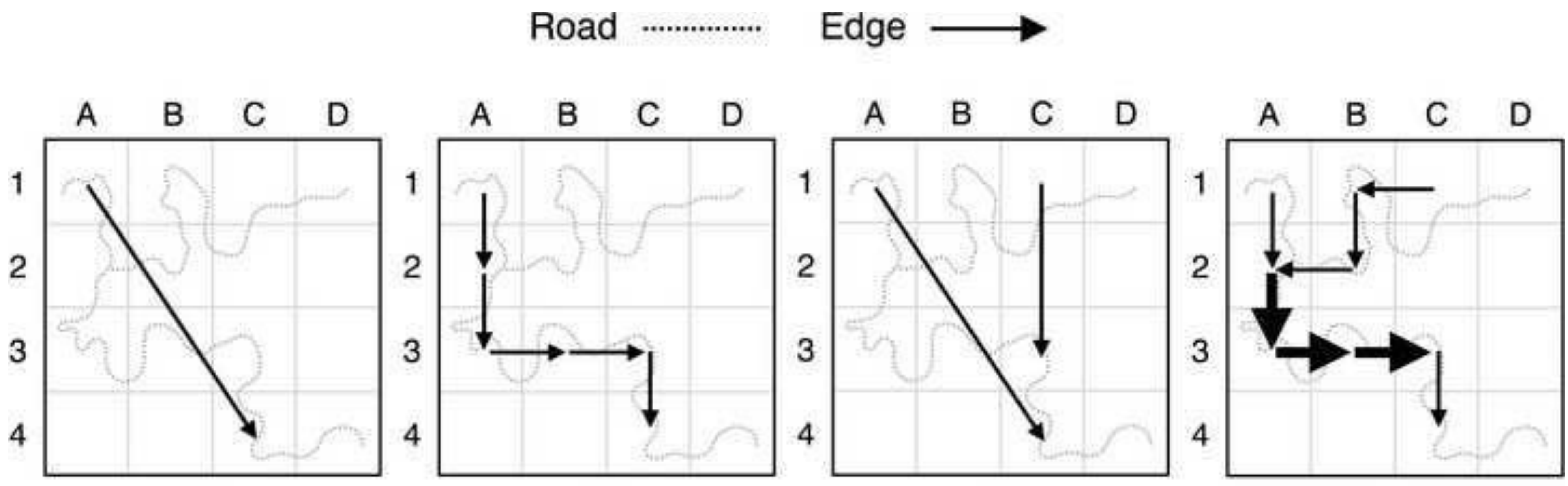


Figure 2
Click here to download high resolution image



(a) (b) (c) (d)

Figure 3

Click here to download high resolution image

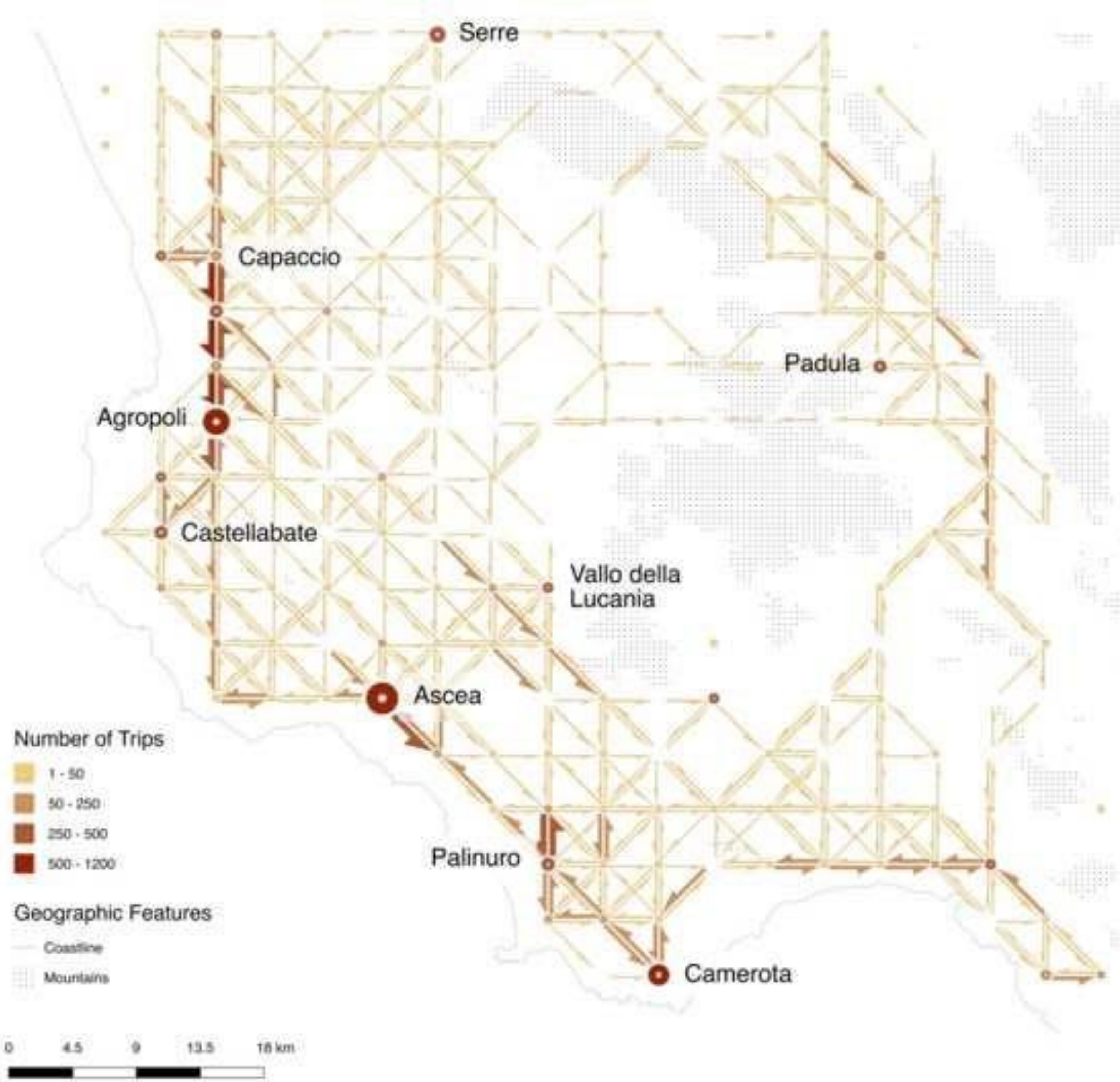


Figure 4
Click here to download high resolution image

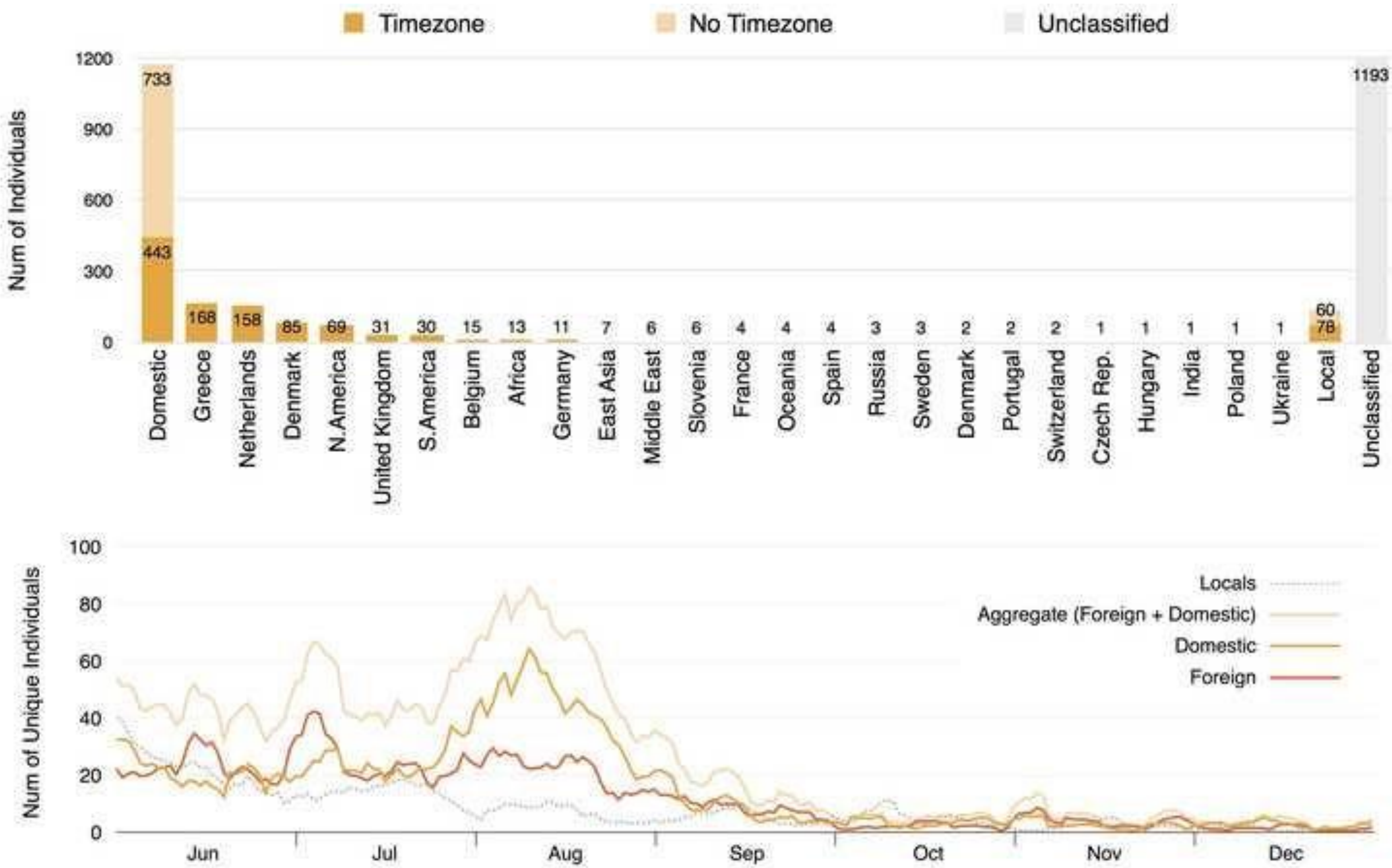
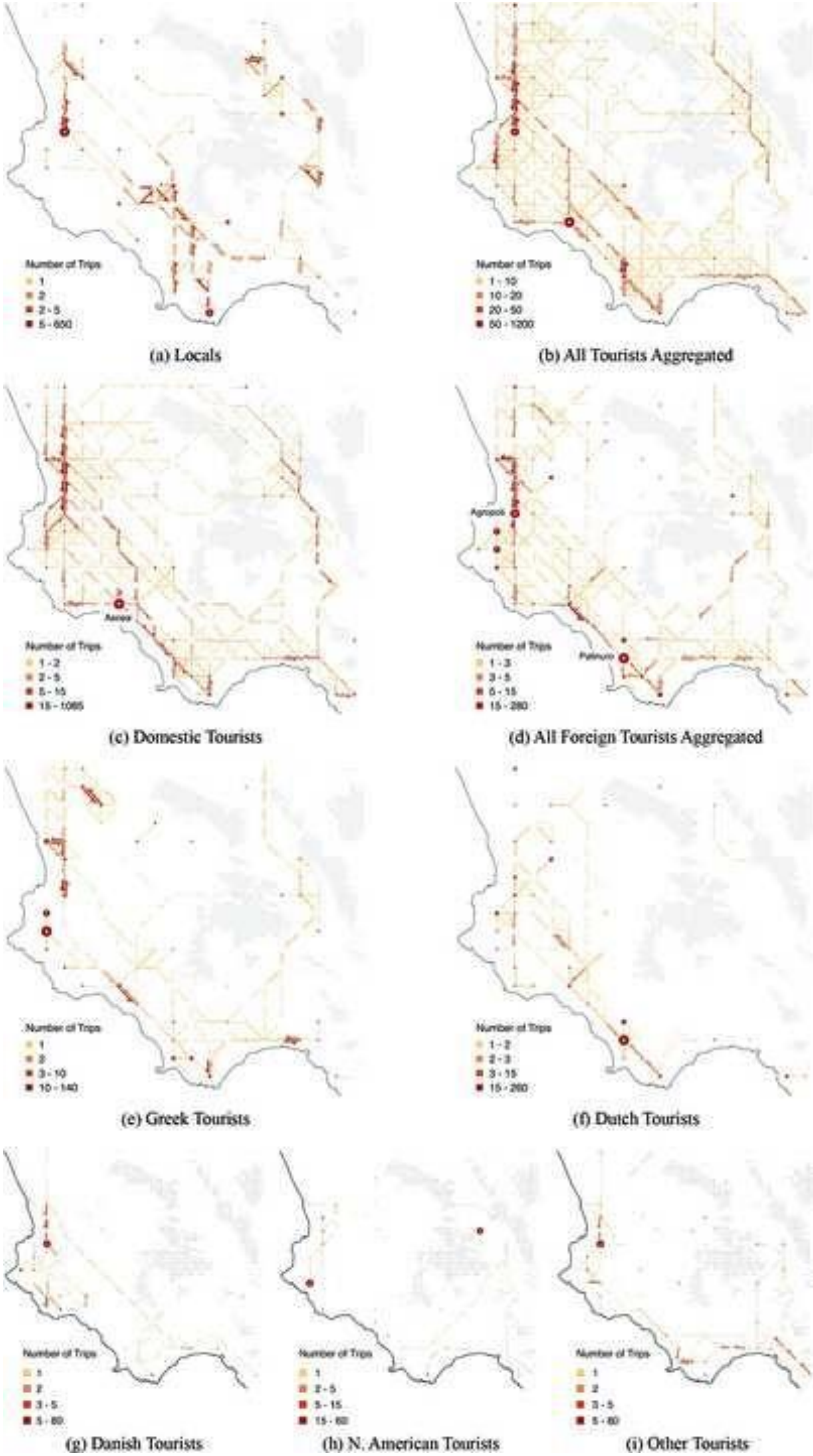


Figure 5

Click here to download high resolution image



(a) Locals

(b) All Tourists Aggregated

(c) Domestic Tourists

(d) All Foreign Tourists Aggregated

(e) Greek Tourists

(f) Dutch Tourists

(g) Danish Tourists

(h) N. American Tourists

(i) Other Tourists

Figure 6
Click here to download high resolution image

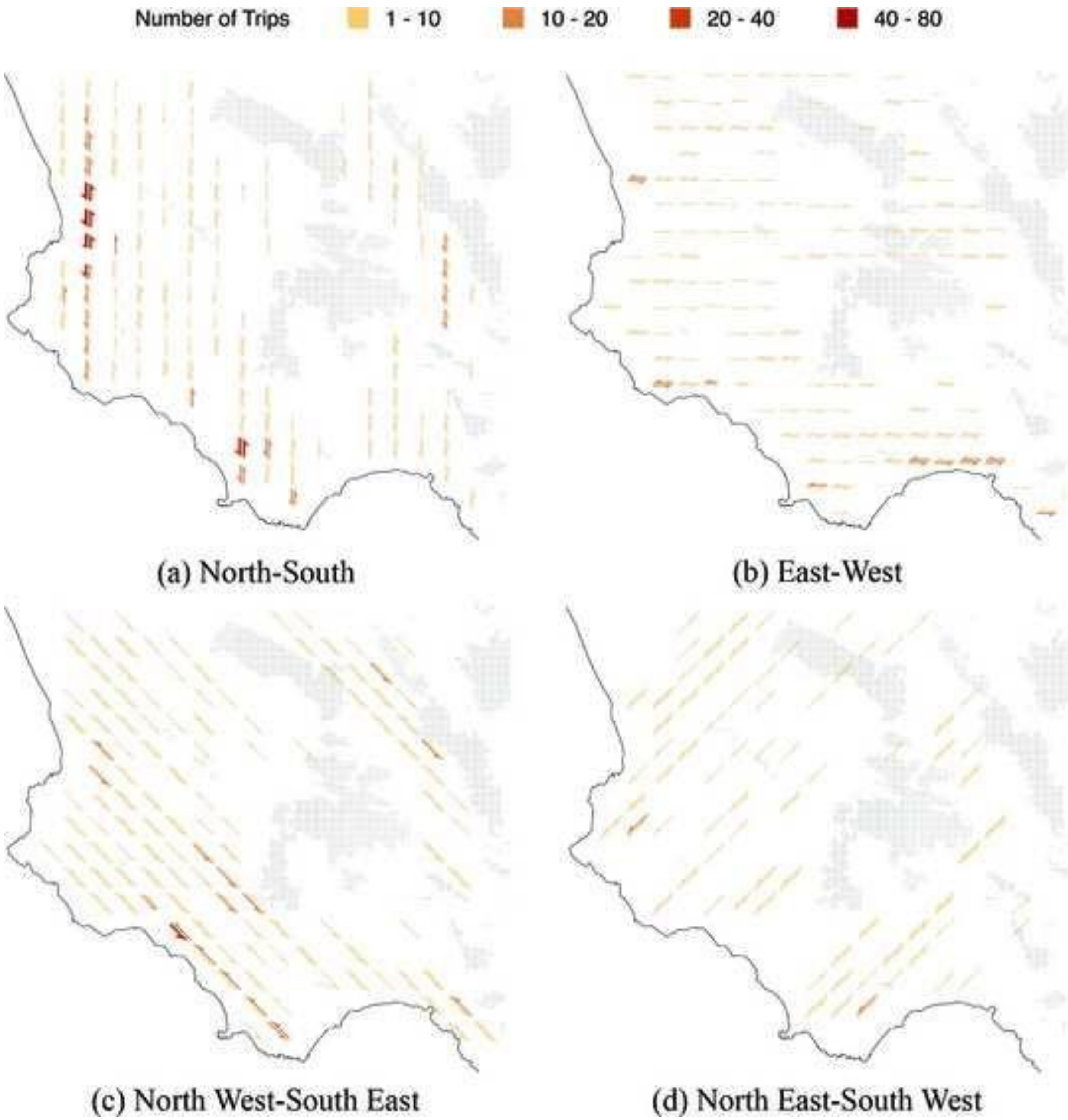


Figure 7

Click here to download high resolution image

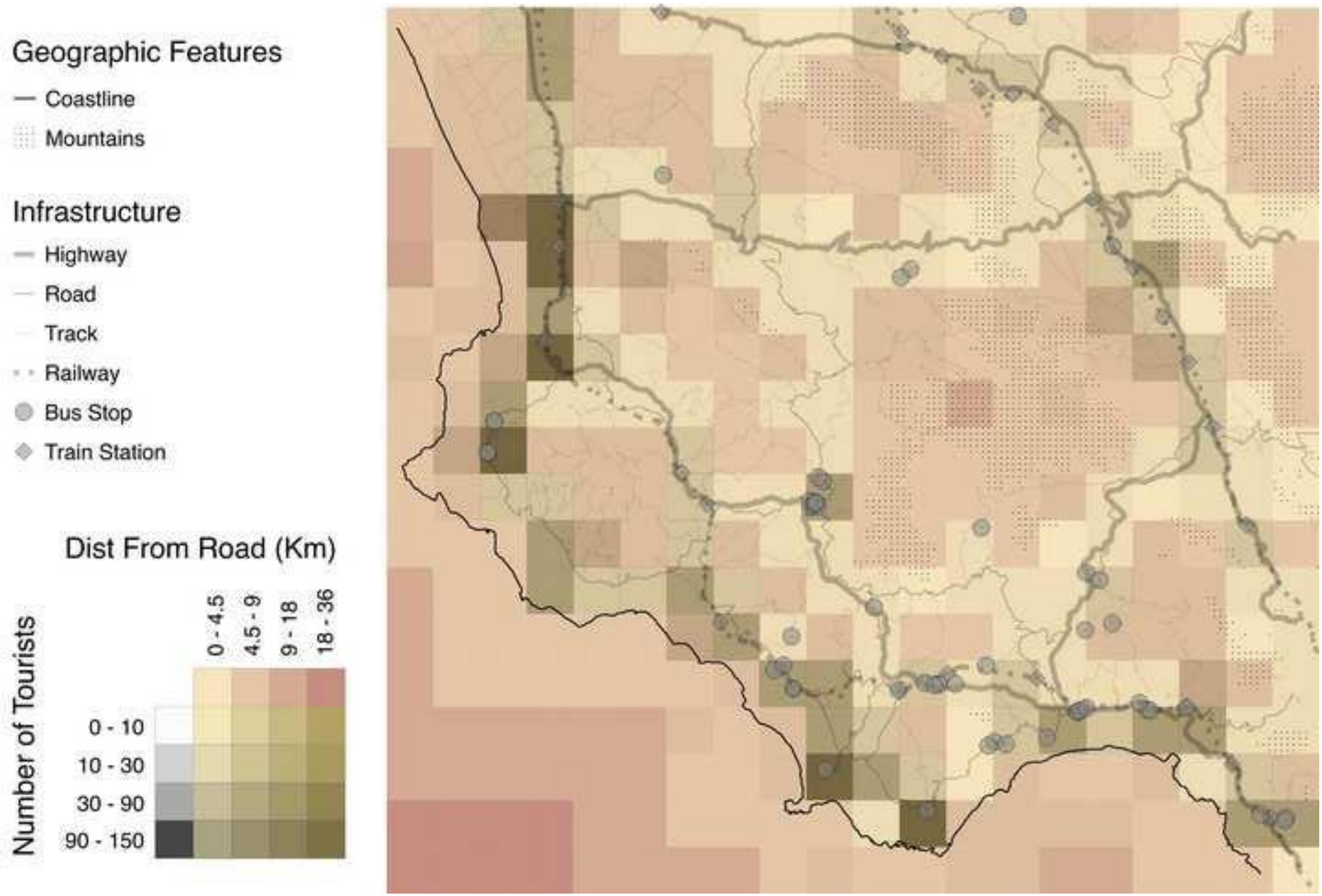
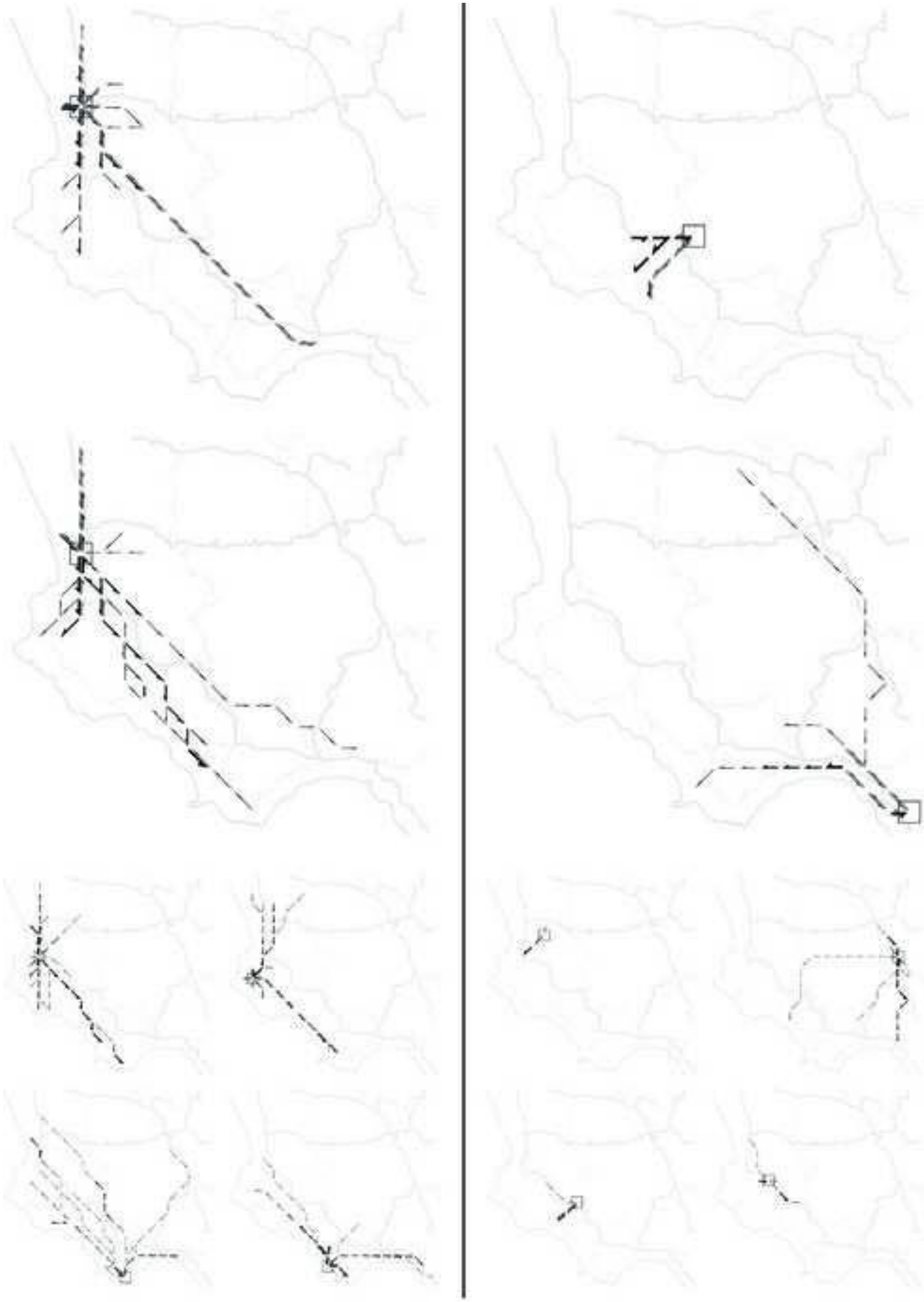


Figure 8
Click here to download high resolution image



(a) Most Popular Locations

(b) Other Locations

***Author Biography**

Alvin Chua
PhD Candidate
Research [x] Design, Department of Architecture, KU Leuven, Belgium

Loris Servillo
Post-Doctoral Researcher
Department of Architecture, KU Leuven, Belgium

Ernesto Marcheggiani
Professor
Division of Forest, Nature and Landscape, KU Leuven, Belgium

Andrew Vande Moere
Professor
Research [x] Design, Department of Architecture, KU Leuven, Belgium

***Author Photo (to accompany biography)**

Alvin Chua

Loris Servillo

Ernesto Marcheggiani

Andrew Vande Moere