Case study

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

Alvin Chua [a, *], Loris Servillo [a], Ernesto Marcheggiani [b, c], Andrew Vande Moere [a]

[a] Department of Architecture, University of Leuven, Belgium
[b] Department of Earth and Environmental Sciences, University of Leuven, Belgium
[c] Department D3A, University of Marche, Italy

## HIGHLIGHTS

- We introduce a novel approach to tourist flow analysis based on geotagged social media data.
- This approach is useful for studying the spatial, temporal and demographic features of tourist flows.
- We show the advantage of our approach through an analysis of tourist movement in South 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 flows 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.

© 2016 Elsevier Ltd. All rights reserved.

## 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, insights 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

* Corresponding author.
E-mail addresses: alvin.chua@asro.kuleuven.be (A. Chua), loris.servillo@asro.kuleuven.be (L. Servillo), e.marcheggiani@univpm.it (E. Marcheggiani), andrew.vandemoere@asro.kuleuven.be (A.V. Moere).

(Christaller, 1964; Mansfeld, 1990). From an operational 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, 2010) 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 movement of tourists 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[1] funded by European and state agencies[2] 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 ha. 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). 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 to broadly understand 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 visited. 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 criteria to determine the quality of data we required.

### 2.2. Limitations with existing data

To 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,

---