Analysis of Tourism Dynamics and Special Events through Mobile Phone Metadata

Yan Leng
MIT Media Lab
Cambridge, MA
yleng@mit.edu

Alejandro Noriega
MIT Media Lab
Cambridge, MA
noriega@mit.edu

Alex 'Sandy' Pentland
MIT Media Lab
Cambridge, MA
pentland@mit.edu

Ira Winder
MIT Media Lab
Cambridge, MA
jiw@mit.edu

Nina Lutz
MIT Media Lab
Cambridge, MA
nlutz@mit.edu

Luis Alonso
MIT Media Lab
Cambridge, MA
alonsolp@mit.edu

ABSTRACT
Tourism has been an increasingly important factor in global economy, society and environment, accounting for a significant share of GDP and labor force. Policy and research on tourism traditionally rely on surveys and economic datasets, which are based on small samples and depict tourism dynamics at low spatial and temporal granularity. Anonymous call detail records (CDRs) are a novel source of data, showing enormous potential in areas of high societal value: such as epidemics, poverty, and urban development. This study demonstrates the added value of using CDRs for the formulation, analysis and evaluation of tourism strategies, at the national and local levels. In the context of the European country of Andorra, we use CDRs to evaluate marketing strategies in tourism, understand tourists' experiences, and evaluate revenues and externalities generated by touristic events. We do this by extracting novel indicators in high spatial and temporal resolutions, such as tourist flows per country of origin, flows of new tourists, tourist revisits, tourist externalities on transportation congestion, spatial distribution, economic impact, and profiling of tourist interests. We exemplify the use of these indicators for the planning and evaluation of high impact touristic events, such as cultural festivals and sports competitions. [1]

Keywords
Tourism planning; Mobile phone records; Data mining; Events analytics; Indicators

1. INTRODUCTION
Tourism has been an increasingly important factor in global economy, society and environment, accounting for a significant share of GDP and labor force [23, 26]. Economists and governments have been attempting to understand the contribution and impact of tourism to economy for years [5, 24]. The growing importance of event tourism, an applied field devoted to understanding and improving tourism

---

[1]This paper is a draft for submission for Data for Good Exchange (2016). The improvements in synthesis, analysis and writing are yet to be implemented throughout.

through events, have resulted in substantial global competition and generated values for both public and private sectors [17]. Research on events and tourism planning, operation and marketing have grown exponentially since 2008, which involves accommodation, attraction, transport and ancillary services[18]. The success of tourism events expand the potential of tourism and the capacity of the tourism area to step beyond a narrow leisure-based tourism [18].

Most tourism and events analytics literatures are based on surveys, interviews, observations or focus group studies [18, 13, 6, 14]. Pettersson (2009) [22] studied the spatial and temporal nature of event experiences through interviews, participant observation and photography at a major sporting event in Sweden, contributing to a better understanding of how visitors interact with the event setting and with each other, and help build theory on experiences. Li (2008) [20] conducted two-phase online surveys with six-month interval to analyze first- time and repeat visitors via demographic characteristics, travel planning behavior, pre and post-trip congruency in travel activity preferences, and post-trip evaluation on first-time and repeated visitors. However, these studies are small in size, unrepresentative, low in spatial-temporal resolutions, and unable to measure travel experiences.

In recent years, big data, mobile sensors and social media, generated many incredible opportunities for its supposed capacity to provide answers for questions related to travelers' behaviors and experiences [4]. Applying big data in tourism has the following advantages: larger reliability than self-reported data and intentions; easier to cross-reference with other data based on the geo-spatial information; more knowledge about the industry's target market produced by the customers themselves. Birenboim (2016) [9] tracked 25 students with an mobile application, which triggers up to three micro-surveys an hour for subjective evaluation of momentary sense of crowdedness and sense of security. Wood (2013) [28] combined field investigation and Flickr data, showing that Flickr data can be used to understand what elements of nature attract tourists and how this would alter visiting behaviors. Bassolas (2016) [7] analyzed the attractiveness of tourists sites using geopositioned twitter.

Large-scale geo-located Information and Communication Tech-

nology, such as Call Detail Records collected from mobile phones, created new ways to understand the interconnectedness among travelers, social events, urban infrastructures [7]. It has been the most salient source of information, eliciting large-scale patterns of human mobility [11]. CDRs have been applied various fields, such as mobility modeling [3], epidemics [15], disaster response [10]. However, this large-scale and longitudinal data source has seldom been used in tourism analytics. In this study, we demonstrate how to utilize mobile phone data to evaluate tourism marketing strategies, understand tourists' experiences, evaluate revenues and externalities generated by tourism.

This paper contributes and adds value to tourism planning and event analytics by novel indicators extracted from mobile phone records. With higher spatial-temporal resolutions comparing with self-reported surveys, CDRs reveals more insights and improved knowledge into various aspects of tourism industry. We extract various indicators, including tourists flows per country of origin, dynamics of new tourists, flows of revisits, tourists externalities due to congestion, spatial distribution, economic impacts and profiling of tourists interests. These indicators help other industries to enhance services they offer and the management. The study can be replicated in other regions and sectors aiding better decision-makings. In the context of Andorra, we use CDRs to evaluate marketing strategies in tourism, understand tourists' experiences, and evaluate revenues and externalities generated by touristic events. We exemplify the use of these indicators for the planning and evaluation of high impact touristic events in 2015, such as cultural festivals and sports competitions.

The remaining of the paper is structured as follows: Section 2 introduces the context of Andorra and the data used in this study. Section 4 describes the novel indicators extracted mobile phone records. Section 5 summaries the performances of events. Finally, section 6 concludes this study and discusses future works.

2. ANDORRA AND DATA
In this section, we briefly introduce the context of Andorra and the data used to conduct the study.

2.1 Andorra
Andorra is a small country in Europe covering an area of 468 km^2, situated near France and Spain, as shown in Figure 1. The population of Andorra is 85,000, with an annual of 10.2 million international visitors [12]. Andorra is a country heavily relying on tourism and famous for skiing and duty-free shopping, with 80% of GDP comes from tourism [1].

2.2 Data
The main data source we use, as stated in section ??, is Call Detail Records (CDRs). It is collected by Andorra telecom, originally for billing purposes. It contains metadata about call, Short Message Service (SMS) and data communications, with information on longitude, latitude, timestamp of the transaction, registry country, phone type, i.e., as shown in Figure 2. The CDR data used in this study was collected over a period of two years from 2014 to 2016. The distribution of cell towers are shown in Figure 3.

Figure 1: Map of Andorra

Figure 2: A snapshot of Call Detail Records

Another dataset coupled with CDR is the road network, enabling the mapping from the cell-tower-based Origin-Destination matrix observed from mobile phone records to traffic flows on road links. We acquired the GIS shapefiles of road network from the Andorra transportation department.

3. VISUALIZATION OF TOURISM
To understand how tourists and locals interact with urban environment - transportation and commercial system - we develop a visualization platform to leverage and cross-reference different geo-spatial data. This provides an integral tool for policy makers to observe visual analytics, uncover hidden patterns and develop actionable insights. Fur-
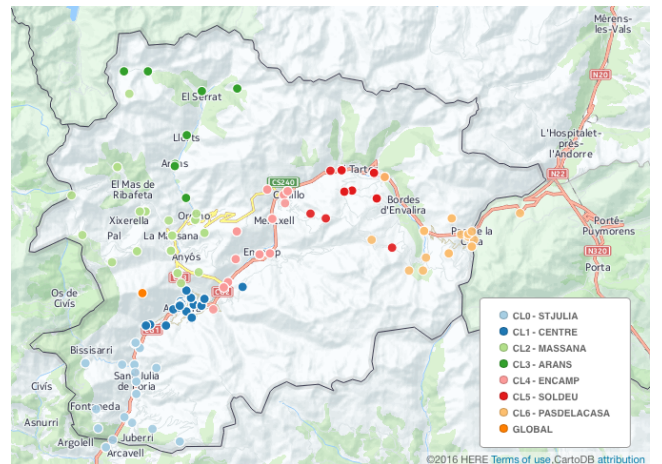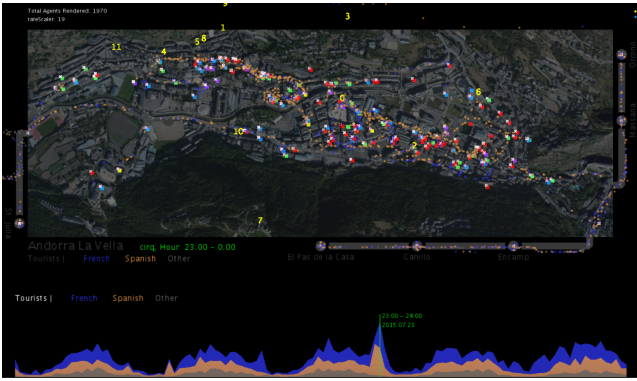

Figure 3: Cell tower distribution in Andorra

Each purple and orange circle represents a French and a Spanish traveler respectively. The colored squares stand for different Points of Interests (POIs).

Figure 4: Visualization of tourists dynamics

Figure 5: Andorra road network

thermore, it make big data accessible and digestiable to narrate, engage and educate a larger audience. Figure 4 is a visualization of an event in 2015, Cirque De Soleil.

4. INDICATORS MINED FROM CDRS

This section aims to describe indicators mined from CDRs, which enable decision-makers and planners to gain deep and multi-dimensional understanding in tourism performance. These indicators cover various aspects, including marketing, travel experiences, economic impacts and externalities.

4.1 Segmented tourists flows

Tourist flow is a traditional indicator in tourism industry, traditionally estimated from survey. It can be measured by tourist volumes or tourist-days, which can be segmented based on country of origins directly observed from Call Detail Records. Visitors from different countries reveal different visitatoin patterns, such as the gathering regions, length of stays, interested events. Understanding the patterns across different market segmentations enables businesses to adapt offers to increase the attraction to tourists. In the context of Andorra, the main countries include Spain, France, Russia, Belgium, Portugal, and Netherlands. Spanish and French take up 90% of the total number of tourists.

4.2 New tourists and repeated tourists

Marketing strategies and the performance events can be measured by new tourists and repeated tourists. Specifically, new tourist volumes reflect the marketing strategy prior to the events. Repeated tourist volumes reclect travel experiences during the event [19, 25]. Correlating repeated visit rates with other factors, such as congestions, implies whether these factors are significant for travel experiences.

New tourists are calculated as the tourists that do not visit prior to the event period. Repeated tourists are defined as tourists that revisit after the event.

4.3 Spatial distribution

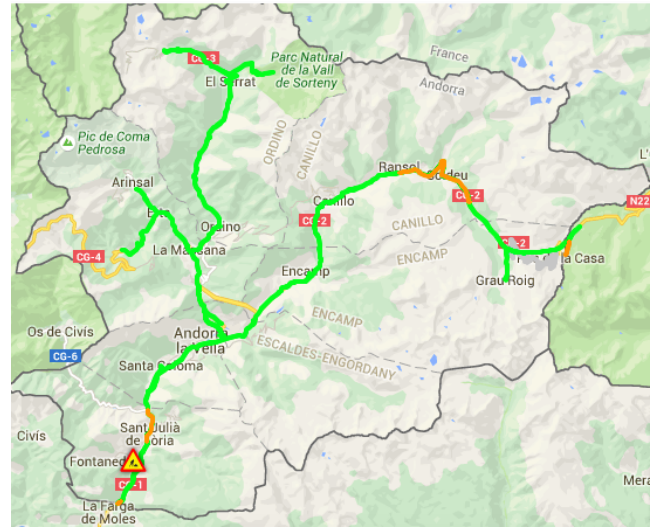Spatial distribution of tourists measure the spatial impact of the event during the event period. Success event generates wider impact, distributing revenues across the nation. It can be measured seperately across daytime and night since activities are different during these two periods.

Specifically for hospitality, understanding the spatial distribution at night based on origin countries enables hotels managers to pinpoint countries to focus marketing.

4.4 Congestion

Congestion is the externality generated by tourism that adversely impact local communitie, travelers' experiences and environmental system. Popular events, seemingly sucessful in attractiveness, are exposed to higher risk of congestion, which, on the other hand, become an disadvantage. Severe congestion during event indicates the need for improving tourist demand management or providing more services.

Mobile phone data provides a dynamic monitoring of transportation system performance. The ability to track traffic flows at a granular spatio-temporal scale provides deeper understanding of tourists' travel patterns, enabling better planning and management.

We propose a three-step method to estimate traffic flow. First, we compute the tower-to-tower Origin-Destination (O-D) matrix by aggregating the individual movements between cell towers. It is important to estimate traffic flows across temporal and spatial dimensions due to the heterogeneous variability. Second, the O-D pairs are assigned to road links, as shown in Figure 5. The last step is to scale the aggregated movements to actual vehicle trips using traffic counts as the ground truth.

4.5 Revenues

Understanding the spending patterns at individual level and aggregated level enables businesses to promote offers to concentrate marketing efforts. It also helps businesses to indentify the potential markets for growth. Under the situation that no data is directly related to tourists' spendings, we come up with an approach to proxy the revenues. Revenues are positively correlated with the disposable income

of tourists, which can, to some degree, be proxied by the price of the phone being used.

The brand, model, operation system of the phone can be obtained with IMEI-TAC code of the records. The price of the phone are scraped from Amazon API based on the above information. The phone type distribution across nations are shown in Figure 6.

Note that accurate revenues and individual spendings generated from events can be measured by banking transactions [27] if available.
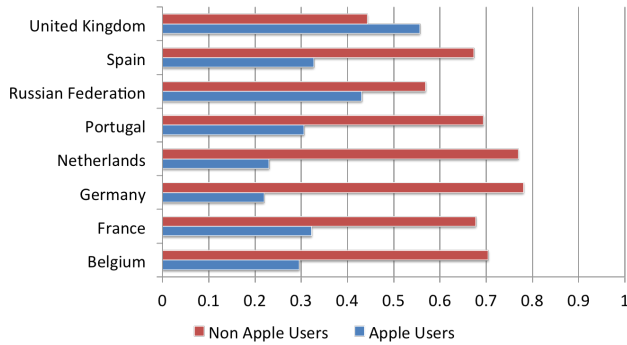

Figure 6: Phone type distribution across country

4.6 Tourists interests profiles

The understanding of tourists' interests offer businesses and tourism department a series of tactical and strategic recommendations on deicision-makings. As detailed footprints are tracked, the behavioral patterns for each interest category increase the knowledge of the industry's target market. This enables businesses to promote various service bundles to retain and attract customers.

Personal interests are inferred in two steps. We first integrate the geo-location of cell towers and Point of Interests (POIs), tagging each cell towers with the potential activities individuals may perform nearby. Second, we aggregate the cell towers each individual connect to. With this, we obtain a distribution for activities individuals are interested in.

5. CASE STUDY

In this section, we use the proposed indicators in section 4 to showcase a comprehensive assessment of the high-impact events based on CDRs, extending beyond traditional indicators learned from surveys.
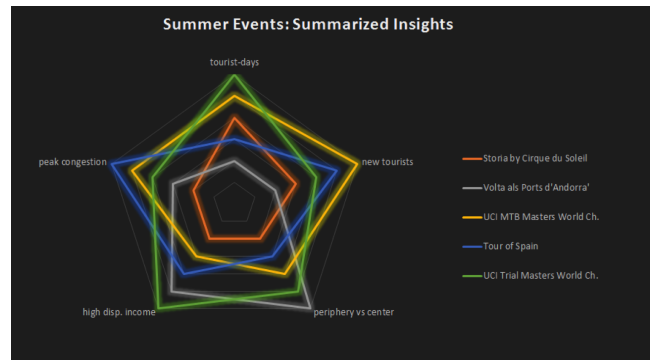
5.1 Summer events analytics

Figure 7 shows the performances of five summer events on a radar chart, based on tourist days, new tourists, spatial distribution, income profiles and peak congestion. We target the following summer events:

- Storia by Cirque du Soleil: 2015/07/15 - 2015/07/20.
- Volta als Ports d'Andorra: 2015/07/12.
- MTB Masters World Championships: 2015/08/23 - 2015/08/27.
- Tour of Spain: 2015/08/31 - 2015/09/03.

- UCI Trial Masters World Championships: 2015/09/01 - 2015/09/06.

Insights from Figure 7 include:

- UCI Trials Master World Championship was relatively more successful in most dimensions. Planners carried out successful marketing strategies and winned high popularity. Activities were not as attractive to high-income visitors as low-income ones, the former ones may generate more revenues.
- MTB World Master Championships attracted a large volume of tourist and new tourists, indicating the popularity and success in marketing strategies. Tourists were centralized in central area and caused severe congestion, indicating more efforts in: 1) marketing suburban areas; 2) organize events or develop businesses in attracting high-income visitors; 3) manage travel demand to reduce congestion.

The relative position indicates the relative performance in the five events. Each spoke represents each indicator, with colors standing for events.

Figure 7: Summer events analytics

5.2 Winter events analytics

Tourist flows during events measures the popularities of the events. However, more insights about the characteristics and segmentation of the population dynamics reveal better knowledge about marketing strategy prior to the events and travel experiences during the events. The targeted winter events include:

- Free-ride Junior World Championship: 2015/02/07 - 2015/02/08
- Speed Skiing World Championship: 2015/02/28 - 2015/03/03
- Alpine Skiing European Cup Finals: 2015/03/18 - 2015/3/22
- Total Fight Master of Freestyle: 2015/03/26 - 2015/04/04

As shown in Figure 8, Free-ride Junior World Championship, though less popular, attracted a large number

of new tourists and repeated visits. This indicates its success in not only the marketing, but also the management and organization of the events. On the contrary, though Total Flight Master of Freestyl had high popularity, lowest revisit rate implies poor travel experiences, requiring more analytics on the specific aspects to be improved to retain tourists.

Figure 8: Winter events analytics

5.3 Tourists interests profiles

The interests of tourists across nations are informative in target marketing. As shown in Figure 9, the Spanish and the French are more interested in shopping and nature activities, which take up $\frac{2}{3}$ and $\frac{1}{4}$ respectively. $\frac{2}{3}$ Russian are interested in nature activities, with another $\frac{1}{6}$ interested in cultural activities. $\frac{1}{3}$ Netherlands are interested in nature and shopping activities.
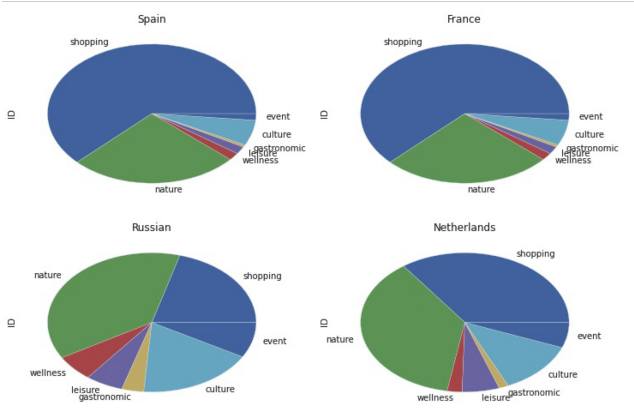

Figure 9: Nationality vs. interests

6. CONCLUDING REMARKS

Growth in leisure travel has become increasingly significant economically, socially, and environmentally, accounting for a significant share of GDP and labor force. Policy and research on tourism traditionally rely on surveys and economic datasets, which are based on small samples and depict tourism dynamics at low spatial and temporal granularity. Anonymous call detail records (CDRs) are a novel source of data, with an enormous potential in areas of high societal value, such as epidemics, poverty, and urban development.

This study demonstrates how to make use of the potential and added value of CDR for the formulation, analysis and evaluation of tourism strategies, at the national and local levels. In the context of the European country of Andorra, we use CDRs to evaluate marketing strategies, understand tourists' experiences, and evaluate revenues and externalities generated by touristic events. We do this by extracting many indicators in high spatial and temporal resolutions, such as tourist flows per country of origin, flows of new tourists, revisitation patterns, tourist externalities on transportation congestion, spatial distribution, economic impact, and profiling of tourist interests. Some of the indicators are traditionally used by tourism industry and others could not be quantified without large-scale and longitudinal data. We exemplify the use of these indicators for the planning and evaluation of high impact touristic events in Andorra, such as cultural festivals and sports competitions.

For future study, we plan to build an overreaching and automatic platform for tourisms and events management. The application of CDR data in tourism is promising in that it can be integrated with other geopositioned data for a comprehensive tourism monitoring and optimization platform, such as WiFi - more granular in movements, social media - momentary feelings and hidden interests, banking transactions - accurate revenue.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] Hoteleria i turisme. 2015.
[2] R. Akcelik. Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. *Australian Road Research*, 21(3), 1991.
[3] F. Alhasoun, A. Almaatouq, K. Greco, R. Campari, A. Alfaris, and C. Ratti. The city browser: Utilizing massive call data to infer city mobility dynamics. In *3rd International Workshop on Urban Computing (UrbComp 2014). UrbComp: New York, NY*, 2014.
[4] R. Baggio. Big data, business intelligence and tourism: a brief analysis of the literature.
[5] J. Balaguer and M. Cantavella-Jorda. Tourism as a long-run economic growth factor: the spanish case. *Applied economics*, 34(7):877–884, 2002.
[6] J. Balaguer and M. C. Jordá. Tourism as a long-run economic growth factor: the spanish case. *Working papers= Documentos de trabajo: Serie EC (Instituto Valenciano de Investigaciones Económicas)*, (10):1, 2000.
[7] A. Bassolas, M. Lenormand, A. Tugores, B. Gonçalves, and J. J. Ramasco. Touristic site attractiveness seen through twitter. *EPJ Data*

Science, 5(1):1, 2016.

[8] M. Ben-Akiva and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, pages 5–33. Springer, 1999.

[9] A. Birenboim. New approaches to the study of tourist experiences in time and space. *Tourism Geographies*, 18(1):9–17, 2016.

[10] V. D. Blondel, A. Decuyper, and G. Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):1, 2015.

[11] A. P. Brian Wheatman, Alejandro Noriega. Electricity demand and population dynamics prediction from mobile phone metadata. 2016.

[12] Cia.gov. Cia world factbook entry: Andorra. 2012.

[13] P. Crowther, J. Bostock, and J. Perry. Review of established methods in event research. *Event Management*, 19(1):93–107, 2015.

[14] A. Dupeyras and N. MacCallum. Indicators for measuring competitiveness in tourism. 2013.

[15] E. Frıas-Martınez, G. Williamson, and V. Frıas-Martınez. Simulation of epidemic spread using cell-phone call data: H1n1 case study.

[16] D. Getz. Event tourism: Definition, evolution, and research. *Tourism management*, 29(3):403–428, 2008.

[17] D. Getz and S. J. Page. *Event studies: Theory, research and policy for planned events*. Routledge, 2016.

[18] D. Getz and S. J. Page. Progress and prospects for event tourism research. *Tourism Management*, 52:593 – 631, 2016.

[19] J. Y. Lee. *Investigating the effect of festival visitors' emotional experiences on satisfaction, psychological commitment, and loyalty*. PhD thesis, Texas A&M University, 2009.

[20] X. R. Li, C.-K. Cheng, H. Kim, and J. F. Petrick. A systematic comparison of first-time and repeat visitors via a two-phase online survey. *Tourism Management*, 29(2):278–293, 2008.

[21] A. Lima, R. Stanojevic, D. Papagiannaki, P. Rodriguez, and M. C. González. Understanding individual routing behaviour. *Journal of The Royal Society Interface*, 13(116):20160021, 2016.

[22] R. Pettersson and D. Getz. Event experiences in time and space: a study of visitors to the 2007 world alpine ski championships in åre, sweden. *Scandinavian Journal of Hospitality and Tourism*, 9(2-3):308–326, 2009.

[23] S. Proença and E. Soukiazis. Tourism as an economic growth factor: a case study for southern european countries. *Tourism Economics*, 14(4):791–806, 2008.

[24] H. Song, L. Dwyer, G. Li, and Z. Cao. Tourism economics research: A review and assessment. *Annals of Tourism Research*, 39(3):1653–1682, 2012.

[25] T. Thiumsak and A. Ruangkanjanases. Factors influencing international visitors to revisit bangkok, thailand.

[26] U. N. W. T. O. UNWTO. World tourism barometer. 2015.

[27] A. B. R. G. e. Victor Sust, Enrique Illera. Big data and tourism: New indicators for tourism management.

[28] S. A. Wood, A. D. Guerry, J. M. Silver, and M. Lacayo. Using social media to quantify nature-based tourism and recreation. *Scientific reports*, 3, 2013.