Understanding tourist multipurpose travel behaviour using Weibo check-ins

Ye, Z*[1], Clarke, G P†[1] and Newing A‡[1]

[1]School of Geography, University of Leeds; Consumer Data Research Centre, University of Leeds and Leeds Institute for Data Analytics, University of Leeds

January 15, 2019

Summary

This paper aims to understand tourist travel behaviours derived from their check-in data on social media. We make use of Sina Weibo check-ins, presenting a novel opportunity to consider in detail the behaviour of one group of tourists (Chinese) in London, a major receiving destination for international tourism. We segment these tourists' based on the activity preferences and spatial mobility patterns extracted from their digital footprints. The check-in data-driven tourist segmentation indicates that different tourist groups have distinctive destination choices and travel patterns. The study concludes that the way shopping activities are included in the multipurpose trips of each tourist segment can be diverse, while a distinct shopping group has been identified.

KEYWORDS: Social media; Weibo check-in data; Multipurpose travel behaviour; Tourist segmentation; London

1. Introduction

Location-based social network (LBSN) data have been widely used to understand the characteristics and movements of populations in a variety of contexts (Salas-Olmedo et al., 2018; Longley and Adnan, 2016). Tourists are one subset of the population for which LBSN data could offer novel insights into tourist flows and tourist travel patterns (Miah et al., 2017; Shoval and Ahas, 2016; Spyrou and Mylonas, 2016; Leung et al., 2013).

Sina Weibo, equivalent to Twitter, is one of the main LBSN platforms in mainland China. It allows users to generate microblogs with check-in at points of interest (POIs). Weibo check-in data are similar to geo-located Tweets and Foursquare check-ins. Weibo check-ins offer a unique opportunity to identify Chinese tourists and extract insights into their activities and mobility (Liu and Wang, 2015). Therefore, after distinguishing tourist users from residents and long-stay visitors, this dataset has valuable potential in understanding overseas Chinese tourist behaviours, including their length of stay, the range and spatial extent of destinations visited and trip purpose.

This research takes Weibo check-in data as an example to address two main questions: what can we learn about tourist travel behaviour from LBSN data? Is it possible to create a tourist segmentation based on individual travel behaviour extracted from LBSN data? We address these questions in relation to Chinese tourists visiting London, for which London is top of the Chinese wish list of travel destination and makes up 93% airport seat capacity from China to the UK (Visit Britain, 2018).

2. Data and methodology

The study makes use of 962 unique POIs in Greater London. Those POIs experienced 'check-in' by at least 10 Weibo users between 1st Jan 2016 and 28th August 2018. A set of 20,233 geotagged check-

* gyzy@leeds.ac.uk
† g.p.clarke@leeds.ac.uk
‡ a.newing@leeds.ac.uk

ins from 6,465 tourist Weibo users were distinguished based on a 20-day timespan criterion (**Figure 1**). Each check-in contains a timestamp and is associated with a named geotagged POI and linked to individual user ID.

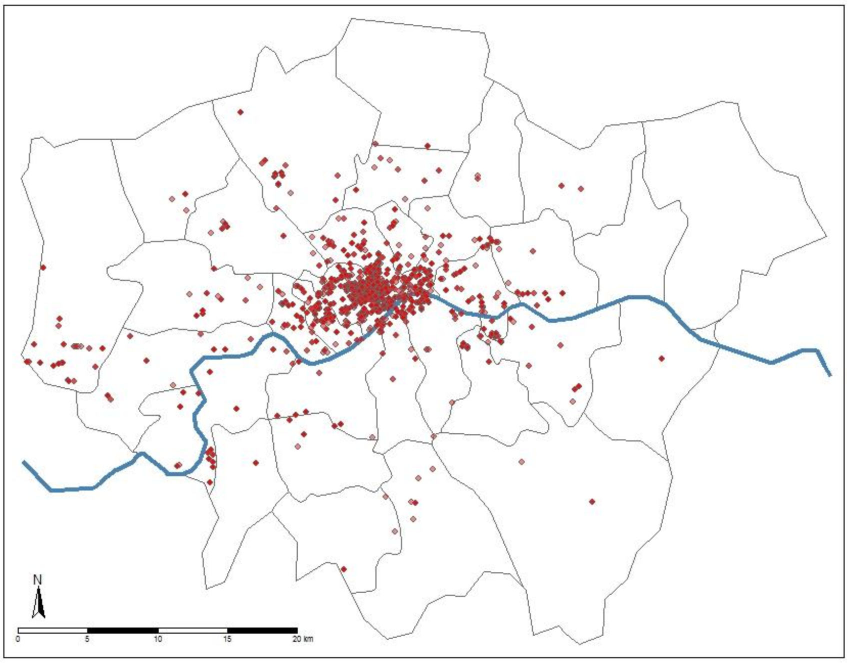

Figure 1 20,233 geotagged Weibo check-ins at 962 unique POIs between 2016-1-1 and 2018-8-28 in Greater London

For further comparison with tourist mobility patterns from other LBSN data, this research associated each Weibo POI with 958 named venues derived from Foursquare venues. In many cases, a single tourist attraction is represented by multiple Foursquare venues and these were combined using DBSCAN clustering such that each POI is represented by a single Foursquare venue. In our final dataset, we have 750 venues frequently visited by our Weibo Tourist users, drawn from the following categories: Arts & Entertainment, College & University, Food, Outdoors & Recreation, Professional & Other Places, Shop & Service, and Travel & Transport.

Thus, Weibo check-in data is enriched to the data structure as shown in **Table 1**. This enables the behaviours and spatial patterns of tourist activity to be explored at an aggregate (venue and attraction) level, or for movement trajectories to be extracted for individual users or groups of users.

Table 1 Data structure of Weibo check-in data after pre-processing

Data field	Example
Check-in ID	151
User ID	1006657733
Check-in date	2018-07-31
Venue ID	4ac518cdf964a520eea520e3
Venue title	Westminster Abbey
Venue detailed category	Church
Venue main category	Professional & Other Places
Venue subcategory	Spiritual Centre
Venue popularity	4751
Attraction name	Westminster Abbey
Latitude/longitude	-0.127356648 / 51.49936

The research procedure and key methods employed in this study are presented in **Figure 2**. This study first depicted Chinese tourist activity hotspots and explored the movements of tourists between attractions at an aggregate level. Using a combination of measurements and indices, we then produced a set of indicators to capture key aspects of individual tourists' travel behaviours based on their daily check-in trajectory. We used K-means to create a Chinese tourist segmentation based on the similarities of their travel modes, activity preferences and mobility patterns. Finally, Latent Dirichlet Allocation (LDA) was adopted to understand the multipurpose travel behaviour of every segment.

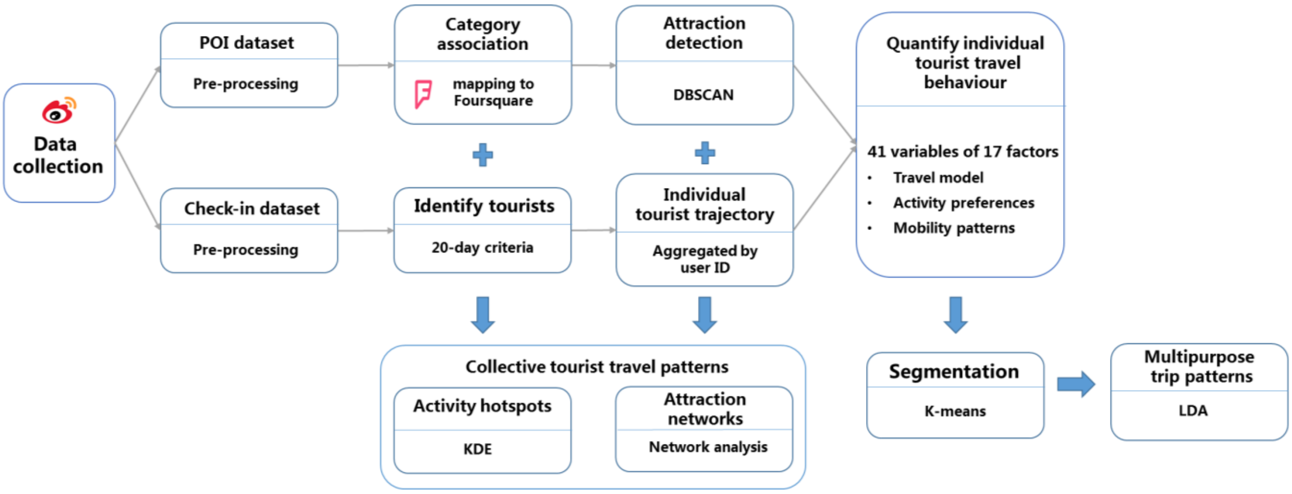

Figure 2 Research procedure and key methods.

3. Collective tourist travel patterns

3.1 Activity hotspots

The main category of each venue refers to different activity type of each check-in. 7 activity types are shown in **Table 2**. The first five activity types are consistent with the top activities for Chinese visitors during their visit to the UK according to VisitBritain (2018).

Table 2 7 activity types of Chinese Weibo tourists and their corresponding categories

Activity type	Venue main category	Typical venue category
Visiting arts and entertainment venues	Arts & Entertainment	Museum, Historic site, Concert hall, Art gallery, Theater
Outdoor sightseeing	Outdoors & Recreation	Park, Palace, Scenic lookout, Bridge, Castle
Visiting monuments/building	Professional & Other Places	Monument/landmark, Church, Library
Dining out	Food	Tea room, Steak house, English restaurant, Burger Joint
Shopping	Shop & Service	Department store, Market, Souvenir shop, shopping mall
Travelling	Travel & Transport	Hotel, Train station, Metro station, Airport
Student activity	College & University	College academic building, College library, College Residence Hall

The first three activities were combined as 'visiting and sightseeing activities'. Kernel Density

Estimation (KDE) was used to visualise the spatial distribution and hotspots of key tourist activities. The KDE adopted a grid size of 400 units to the outline of Greater London to produce smooth images for visiting and sightseeing, shopping, dining out and travelling activity respectively (**Figure 3**). The visiting and sightseeing activity has two significant hotspots centred on both sides of the River Thames near Westminster Bridge, and 5 secondary hotspots centres distributed in locations where important tourist attractions such as the British Museum, Hyde Park, the Sherlock Holmes Museum, National History Museum and Tower of London are located. Shopping activity and dining out activity both have 2 main hotspots areas. They shared one hotspots area near Knightsbridge and the other at Oxford Street and Soho area respectively. Travelling activity has a dispersed and evenly distributed KDE surface because of the widely distributed hotels and transport stations.

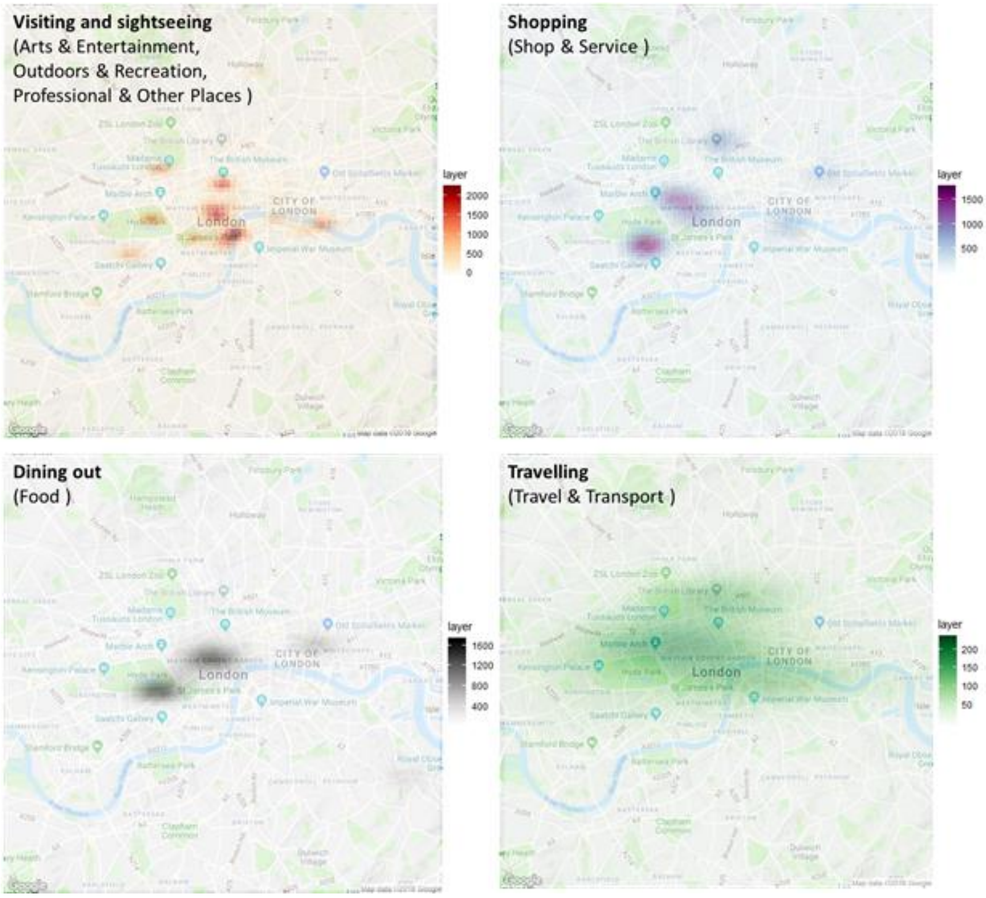

Figure 3 Kernel density estimation (KDE) of different tourist activities: (a) visiting and sightseeing, (b) shopping, (c) dining out, (d) travelling.

3.2 Attraction networks

Tourist behaviours were investigated at the attraction level. The most visited attractions by Weibo tourist users are listed in **Table 3** and compared with the top popular attractions according to the statistic of VisitBritain (2018). It seems that Chinese tourists have stronger preferences for attractions related to British culture, whereas the general visitors (as captured by VisitBritain) are more likely to visit museums, galleries and historic sites.

Table 3 Top attractions in Greater London by Weibo tourists' check-ins, VisitBritain statistics and eigenvector centrality

Weibo tourists' check-ins	VisitBritain (2018)	Attraction centrality
London Heathrow Airport	British Museum	London King's Cross Railway Station

The London Eye	Tate Modern	The London Eye
British Museum	National Gallery	Hyde Park
Hyde Park	Natural and History Museum	Trafalgar Square
Tower Bridge	V&A Museum	British Museum
Buckingham Palace	Science Museum	Buckingham Palace
221b Baker Street	Somerset House	Chinatown
Big Ben	Tower of London	The Sherlock Holmes Museum
Harrods	Tate Britain	Notting Hill Gate Underground Station
National Gallery	National Portrait Gallery	London Heathrow Airport

Network analysis was used to reveal underlying mobility patterns which link different attractions. We used eigenvector centrality to understand the strength of ties between attractions. We identified that 22 attractions act as key nodes, frequently visited by tourists as part of multi-purpose trips involving multiple attractions (**Figure 4)**, The attractions with top centrality also feature in **Table 3**, which suggests that these attractions are the core part of attraction networks and functioned as transit during tourist multi-purposed daily trips.

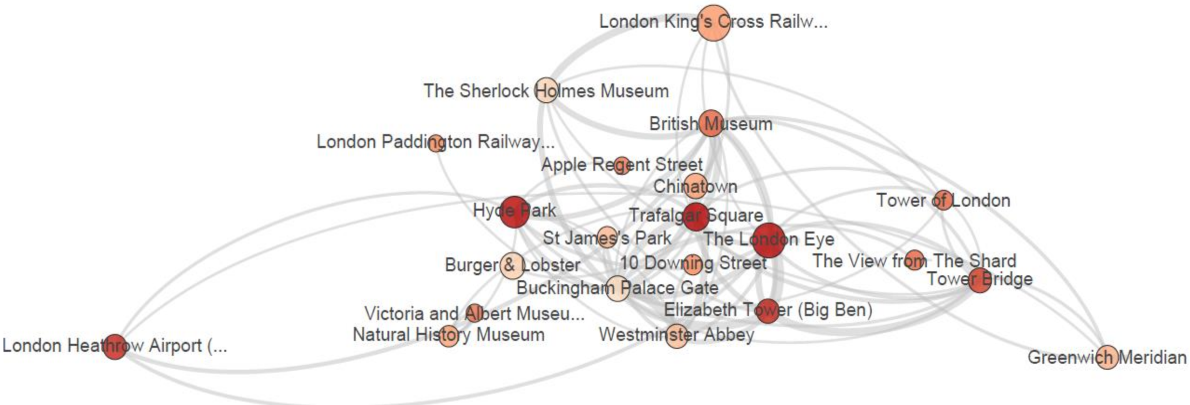

Figure 4 Core attraction networks based on Chinese tourists' daily check-ins trajectories.

4. Quantify individual tourist travel behaviour

Tourist travel behaviour at the collective level helps us to obtain an overall view of how Chinese tourists travel in Greater London. However, every tourist has different travel movements and activities which reflects how they experience London differently. Therefore, the second stage of this study introduced a series of variables which can be used to understand individual tourist travel behaviours. 41 variables are used to quantify individual tourist travel behaviour from different travel modes, activity preferences and mobility patterns (**Table 4**).

Table 4. Description of variables used to capture tourist behaviours

	Domain	Variable	description
1	Travel modes	Length of stay	Number of days between first and last check-in
2		Number of trips	Number of days having check-ins
3		Number of stops	Number of check-ins
4		Number of different attractions	Number of check-ins at different attractions
5-11	Activity preferences	Percentage (7 different activities)	Percentage of one activity to all 7 activities
12-18		main activity percentage (7 different activities)	Percentage of the number of days that one activity as the main activity to the number of trips

19-25		Diversity (7 different activities)	Number of venues in each activity
26-32		Popularity (7 different activities)	Sum of the popularity of visited places of each activity
33		Multipurpose degree	Average number of activities per trip
34	Mobility patterns	mean transit attractions	Mean transit of attractions per trip
35		Periodic behaviour	Returning probability
36		mean travel distance	Mean travel distance per trip
37		mean travel placement	Mean distance between stops
38		Dispersion of footprints	Area of Standard Ellipses Deviation (SED)
39		Distribution of footprints	The eccentricity of Standard Ellipses Deviation (SED)
40		Total weight of routes	The total weight of a tourist's travel route according to attraction networks
41		Total degree of attractions	The total degree of a tourist's travel route according to attraction networks

5. Tourist segmentation and their multipurpose trip understanding

The 41 variables outlined in **Table 4** were used to segment tourists into distinct groups based only on their daily movement records. This study first used k-means to create a segmentation of 1,171 Weibo tourists who have daily routes consisting more than 3 stops, then used LDA algorithm to further understand the multipurpose trip characteristics of tourists within each classification group. LDA has been widely used to identify topics from documents and calculates the proportion of different topics in each document by examining word distributions in the documents (Blei et al., 2003). In our work, we treat the daily trajectory of each Weibo tourist user as a "document" and the subcategory of each place as a "word". The "topics" are hidden check-in patterns of each tourist.

The 41 variables were used in k-means to classify these tourists into 5 clusters which can be described as follows:

Cluster 1 (31.4%) is a group of tourists that only have a high proportion of check-ins at Arts & Entertainment venues. They have a longer length of stay but few of their daily trips are shared on Weibo. The topic modelling results of Cluster 1 unveil that a subgroup of these tourists is likely to include shopping activities at department stores, markets and souvenir shops as important parts of their multipurpose trips.

Cluster 2 (25.3%) consists of tourists who mainly visit very popular outdoor sightseeing attractions but seldom have any shopping activities. Their footprints are widely dispersed during their short stays, but they do not have very busy daily schedules, which is in line with their large mean displacement between attractions. The result of LDA show homogenous results of this group of tourists: they seldom have any shopping related check-ins except department store occasionally appearing. There is a clear subgroup of tourists who prefer churches, while another subgroup has more check-ins generated at hotels.

Cluster 3 (12.1%) is a group of tourists who have a very strong preference for dining out and relatively lower propensity to undertake sightseeing activities. They do not stay long in London, but they tend to share most of their trips on Weibo. They enjoy shopping a lot but it is rare as their main trip purpose and they only visit the most popular department store shopping venues, souvenir shops and markets. They travel among the core part of the attraction networks but take rather relaxed trips (fewer visited attractions per trip). By further using LDA for topic modelling, we found that the tourists in this group are heterogeneous. Apart from a distinct subgroup of tourists which have an apparent fondness for performing arts, a clear difference of dining choices also can be identified: a subgroup prefers Asian restaurants while another subgroup chooses more western food at burger joints and steak houses.

Cluster 4 (10%) is a highly homogenous group characterised by a very high proportion of check-ins at College & University venues with an apparently periodic behaviour. They also enjoy outdoor sightseeing and do shopping at diverse but popular places such as department stores, markets, electronic stores, souvenir shops, clothing stores, and bookshops. Their footprints are the least evenly distributed and shown at some peripheral areas in the attraction networks including some unpopular Travel & Transport venues.

Cluster 5 (9.9%) is a group with clear shopping activities and also outdoor sightseeing. They check in at a variety of popular shopping venues. They have longer stay in London compared with other clusters but they still travel with rather busy schedules. Thus, they have more multipurpose trips with a high diversity of activities. Their travel routes are very typical but only remain in a small region of London. They enjoy a variety of outdoor attractions and their shopping activities are most likely to occur at department stores. After topic modelling, the tourists in Cluster 5 have very heterogeneous and highly diverse shopping activity at department stores, markets, electronic stores and flea markets.

6. Conclusion

This study demonstrates LBSN data have potential in understanding tourist activities and mobility patterns at both a collective and individual level. Although the collected Weibo tourist check-in data cannot reflect every detail of tourist movements, they provide an alternative approach to statistic data from surveys to obtain insights into tourist travel behaviour. By extracting tourists' travel modes, activity preferences and mobility patterns from Weibo check-in data, this study quantified individual Chinese tourists based on their travel behaviours and segmented into 5 groups. The segmentation results indicate that each Chinese tourist group has distinct travel behaviour, and involves different degrees of shopping activity in their multipurpose trips (including one group with a strong preference for shopping). The study added a spatial dimension of tourist travel behaviour in tourist segmentation and revealed different multipurpose travel choices, especially consumption-related behaviour of each segment.

The current research will extend to other mainstream LBSN data such as Twitter and Flickr to generate further knowledge of tourist consumption-related travel behaviour in London. These prevailing LBSN platforms will contribute larger-scale tourist samples which richer and more accurate knowledge about tourists can be extracted and form a dynamic microgeography of London as a whole. Also, more information can be extracted from LBSN data to enrich and refine individual travel behaviour. For example, textual information can offer insights into tourists' specific activity and related experiences; travel times from the post timestamps can help to better infer activity types. These information has great potential to better understand tourist behaviours in relation to specific segments (e.g. retail) in more detail and therefore can be used to improve planning of those tourism-related services.

References

Blei, D., Jordan, M. and Ng, A.Y. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*. **3**, pp.993–1022.

Leung, D., Law, R., van Hoof, H. and Buhalis, D. 2013. Social Media in Tourism and Hospitality: A Literature Review. *Journal of Travel and Tourism Marketing*. **30**(1–2), pp.3–22.

Liu, X. and Wang, J. 2015. The geography of Weibo. *Environment and Planning A*. **47**(6), pp.1231–1234.

Longley, P. a. and Adnan, M. 2016. Geo-temporal Twitter demographics. *International Journal of Geographical Information Science*. **30**(2), pp.369–389.

Miah, S.J., Vu, H.Q., Gammack, J. and McGrath, M. 2017. A Big Data Analytics Method for Tourist Behaviour Analysis. *Information and Management*. **54**(6), pp.771–785.

Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C. and Gutiérrez, J. 2018. Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*. **66**, pp.13–25.

Shoval, N. and Ahas, R. 2016. The Use of Tracking Technologies in Tourism Research : A Review of the First Decade. *Tourism Geographies*. **6688**(September), pp.1–20.

Spyrou, E. and Mylonas, P. 2016. A survey on Flickr multimedia research challenges. *Engineering Applications of Artificial Intelligence*. **51**, pp.71–91.
Visit Britain 2018. Market and Trade Profile: China. (2018, November), pp.5–39.

Biographies

Zi Ye is a second year PhD student with research interests in retail geography and geospatial data analysis.

Graham Clarke is a Professor of Business Geography with research interests in GIS, urban services, retailing and urban modelling.

Andy Newing is a Lecturer in Retail Geography with research interests in spatial and quantitative analysis for retail location planning, census and neighbourhood analysis and service planning.