Advancing reliability assessment of venue-reference social media data for enhanced domestic tourism development

ORIGINAL RESEARCH

Advancing reliability assessment of venue-reference social media data for enhanced domestic tourism development

Huy Quan Vu[1] · Shah Jahan Miah[2] · Haiyang Xia[3] · Gang Li[4] ·
Birgit Muskat[3] · Rob Law[5]

© The Author(s) 2023

Abstract

The aim of this research is to assess the reliability of venue check-in data. We focus on resident mobility and assess how venue check-in data can provide valuable insights into residents' movement in tourist destinations. Understanding resident's mobility and related activities is necessary for strategic planning and decision making for destination managers. Venue-referenced social media data (VR-SMD) is a new type of data and can provide insights into resident's mobility and behavior. While VR-SMD is available on various social media platforms, their reliability to capture resident movement and activities has not been fully realized. We present an analytical approach to assess and demonstrate how VR-SMD, specifically venue check-in data, can be used to capture resident mobility and activities, with illustrative examples from the United States of America, United Kingdom, and Singapore. Results and analysis indicate that tourism managers and government organizations can exploit VR-SMD for domestic tourism development.

Keywords Mobility tracking · VR-SMD · Venue check-in · Domestic tourism

1 Introduction

As one of the world's most important industries, tourism contributes to the continued growth and development of the global economy (Statista 2020; WTTC 2020a). However, these economic benefits of tourism could severely be affected by major events, such as the disease outbreaks (e.g., SARS in 2003 [Demmler and Ligon 2003] and COVID-19 in 2020 [Assaf et al. 2022]) or the economic crises triggered by war (e.g., Ukraine–Russia conflict in 2022 [Pandey and Kumar 2022]). In times of crisis, domestic tourism becomes the focus of governments for tourism development and economic recovery, as people's ability to travel internationally are often limited (TRA 2020).

Extended author information available on the last page of the article

Understanding resident's mobility in tourist destinations is critical for local tourism development (Orellana et al. 2012; Vu et al. 2015) and transportation management (Chua et al. 2016). Various big data sources have been utilized to track people mobility, such as mobile tracking data from service providers (Xiong et al. 2020; Zhou et al. 2020) and geo-tagged posts from micro-blogging social medial platforms (Huang et al. 2020; Li et al. 2020). While the widespread availability of location-tracking technologies makes their data suitable for this purpose, a common limitation of these technologies is that the data lack contextual information about the activities performed at a location. Such knowledge would provide important insights into the reasons for travel, facilitating the design and implementation of targeted and effective domestic tourism development policies.

Recently, a relatively new form of social media based big data, *venue-referenced social media data* (VR-SMD) (Vu et al. 2020), has been introduced, and it appears to have significant advantages over traditional geo-referenced data sources. These data can provide contextual information about people's visits to specific locations. Venue check-ins generated by mobile applications are an example of VR-SMD, where users can conveniently generate check-ins at the time of their visit. Check-in platforms such as *Foursquare*[1] contain comprehensive databases of venues worldwide, and users can conveniently generate check-ins via mobile applications when visiting each venue (Jin et al. 2016). VR-SMD can include information on visited locations and contextual information regarding people's activities at each location, and meta-data can help infer related activities (e.g., dining for restaurant and shopping for shop visits). While extant research has used venue check-in data (e.g., Luo et al. 2019; Vu et al. 2019, 2020), it remains unclear if it can be used to capture residents' mobility and provide valuable insights for domestic tourism development.

In this paper, we present an analytical approach to assessing the reliability of VR-SMD in the form of venue check-in to capture resident's mobility to assist tourism managers in domestic tourism development. As case demonstration, we use *Foursquare* (https://foursquare.com/) venue check-in data collected in the United States of America (USA), United Kingdom (UK) and Singapore for our experiments. Since *Foursquare* rewards users with points for staying at or very close to a selected venue (Patil et al. 2012), we consider these data to be genuine. We first assess venue check data reliability by comparing them with other benchmark data sets. We then perform exploratory analysis to reveal travel behaviors of residents to demonstrate the benefits of using these data for domestic tourism development. In the context of this study, residents refer to people who are residents or stay in a country for a long period. Their mobility can be associated with daily life and tourism activities in a domestic context.

---

[1] https://foursquare.com/.

2 Literature review

2.1 Mobility tracking in travel behavior analysis

Capturing and understanding people mobility is one central aspect for economy development of countries worldwide, especially in the tourism sector (Shoval and Ahas 2016; Stienmetz and Fesenmaier 2019). Traditional approaches have often relied on surveys and questionnaires or GPS trackers to record participant travel histories (Orellana et al. 2012; Shoval et al. 2011), but these approaches are costly, time-consuming, and limited in the scale of the surveyed area, rendering them inappropriate for large-scale analysis (Zheng et al. 2011). Alternative approaches to research tourist mobility have used entry ticket data from theme parks (Joo et al. 2014), bank card transaction details (Sobolevsky et al. 2015), and Bluetooth (Versichele et al. 2014) and mobile (Raun et al. 2016; Baggio and Scaglione 2018) tracking data. One major drawback of these data is that they usually have restricted access, preventing their wide adoption.

Other large data sources include geo-tagged photos on photo-sharing sites (e.g., Flickr) (Ghermandi et al. 2020; Payntar et al. 2021; Stienmetz and Fesenmaier 2019; Vu et al. 2015) and geo-tagged posts on micro-blogging platforms (e.g., Twitter, Weibo) (Chua et al. 2016; Li et al. 2020). Mobility data can be retrieved from social media platforms using application programming interfaces (APIs). Although location data on these platforms contain raw GPS information, which is convenient to track an individual's movement, they have limited option to provide contextual information about travelers. Most existing studies using such data sources focus mainly on people movements. Insights into their activities at their visited locations are lacking.

2.2 Check-in data for resident activity studies

Recently, VR-SMD have been introduced into tourism research as a new data source to study tourist travel behavior (Vu et al. 2020). According to Vu et al. (2020), these data can take various forms as long as they capture the venue information of the visited location. For instance, review comments on travel-related platforms such as TripAdvisor, Expedia, and Yelp, can be treated as VR-SMD if they are analyzed together with the venue information. A limitation of such platforms is that they cover limited venue types (e.g., restaurants and hotels). Besides, reviews are not necessarily posted at the same time as the visit, these review comments are less suitable for capturing resident movements. VR-SMD in the form of venue check-ins have demonstrated advantages in providing rich information about the travel patterns and activity preferences of tourists (Vu et al. 2019; Luo et al. 2019).

Due to the recent global pandemic by COVID19, venue check-in data has been receiving increasing attention from government and tourism managers as an effective tool to support their decision making. Governments in many countries (e.g., Australia [ACT 2021], United Kingdom [NHS 2020]) have created various check-ins

applications and enforced check-ins for residents when visiting venues to track their movement during the pandemic. The venue check-in applications created by governments and their enforcement policies have a major drawback, because they are usually applied to only a few venue types (e.g., restaurants and hotels) where social distancing is difficult. A wide range of other venue types (e.g., parks, stations, and shops) are not covered. As the pandemic comes to an end, such applications are no longer used by residents to record their travel activities.

Venue check-ins on social media platform (e.g., *Foursquare, Yelp* and *DianPing*), on the contrary, cover a wide range of venue types and are not limited to a specific geographical area or administrative region. User check-ins on social media platforms are voluntary, because users are motivated to establish an image of themselves or connect with friends (Cranshaw et al. 2011). VR-SMD can potentially be a valuable data source to support tourism destination managers and government organizations in understanding not only residents' mobility but also their local activity patterns.

2.3 Summary

This paper makes important contributions to extend existing works: *First*, the majority of prior works focuses mainly on popular tourism context, such as inbound tourism (Vu et al. 2015; Raun et al. 2016; Wörndl et al. 2017; Zheng et al. 2017; Salas-Olmedo et al. 2018; Mou et al. 2020; Ghermandi et al. 2020; Payntar et al. 2021) or outbound tourism (Barchiesi et al. 2015; Vu et al. 2018, 2020). We extent these studies and are among the first to study the mobility behavior of residents from the perspectives of daily life and domestic tourism in their home city and country. *Second*, VR-SMD in the form of venue check-ins have been utilized to study the behavior of tourists at tourism destinations. However, it remains unclear how reliable venue check-in data are in capturing travel behavior of local residents. Thus, the contribution of this paper is on the reliability assessment of VR-SMD in the form of venue check-in. Further, we offer a guidance for practical implications, rather than simply demonstrating their use cases as in the prior works (Luo et al. 2019; Vu et al. 2019, 2020).

3 Methodology

The process to collect VR-SMD in the form of venue check-ins is first presented, followed by the details about our approach to assessing the data's capability in capturing residents' mobility. The analytic procedure to provide insights into daily life and domestic tourism activities is then described.

3.1 Venue check-in data collection

We use *Foursquare* venue check-ins as a VR-SMD source. These data are associated with defined venue types (Foursquare 2020) grouped into 10 major categories: Art & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors

& Recreation, Professional & Other Places, Residence, Shop & Service, and Travel & Transport. Users cannot be identified, or their historical check-in records tracked. However, *Foursquare* has integrated its platform with *Twitter*, so *Foursquare* check-ins appear on Twitter as tweets with check-in information.

Twitter data can be extracted using the platform's API, which provides a streaming function that enables users to collect tweets generated in real time. The locations of collected tweets can be specified by a bounding box whose coordinates are defined by minimum longitude (min_lng), maximum longitude (max_lng), minimum latitude (min_lat), and maximum latitude (max_lat). The bounding box is usually defined to cover the entire geographical area of interest (e.g., city or country). Tweets generated within the bounding box but not belonging to the region of interest can be filtered out based on GPS information. If no bounding box is provided, the function returns random tweets generated worldwide.

Tweets, generated by check-ins are differentiated from ordinary tweets by the keyword "swarmapp," the mobile application to generate check-ins on *Foursquare*. An issue with the *Twitter* streaming function is that returned tweets are randomly selected from all possible tweets generated at a time, with inclusion of all check-in tweets generated by an individual user not guaranteed. Accordingly, we collect data in two stages. We use the streaming function to extract a set of seed check-in tweets, allowing identification of *Twitter* users who also use *Foursquare*'s check-in feature. We then use the function *getUserTimeline* to extract all recent tweets from the timelines of identified users, and the keyword *swarmapp* to filter ordinary tweets from check-in tweets. Details of check-ins (e.g., local date, time, GPS location) and visited venues (e.g., venue and category names) are extracted by querying original *Foursquare* check-ins via a unique link attached to each check-in tweet.

Temporal (date, time) and geographic (GPS location) information contained in check-in data are similar to other data sources, such as geo-tagged posts used in previous studies (Huang et al. 2020; Li et al. 2020). However, venue information (venue category) provides additional contextual information about activities.

3.2 Benchmark assessment data sets

Once, the venue check-in data are collected from a social media platform like Foursquare. We assess their reliability by comparing them with another benchmark data set for resident mobility. Data generated from mobile signals are often deem reliable to capture resident's mobility (Grantz et al. 2020), thanks to the widespread use of mobile phones among the public. Mobile phone location data can be generated passively through routine interaction between the phone and service provider, and they usually consist of GPS location of the device and timestamp and unique identifiers of users. Such raw location data have been employed in several studies on public movement in response to government policies (Zhou et al. 2020; Xiong et al. 2020). These are the same type of data used to study tourist movement in some prior studies (Versichele et al. 2014; Raun et al. 2016). These data are owned by private telecommunication companies and, thus, are often restricted from public access.

Another form of passively generated location data is those generated by smartphone applications with GPS function enabled. These data are freely accessible to public via data portals, which are created by geographical information service providers to support policy development of government organizations (Wellenius et al. 2020; Aktay et al. 2020). An example of such data source is COVID-19 Community Mobility Reports (CCMR)[2] created by Google since the start of COVID-19 pandemic in 2020. These reports contain data representing how visits at different places change over time, which are available by geography and across different categories of places (e.g., retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential). Other data portals with similar characteristics to Google's CCMR are Apple's mobility trend report[3] and Citymapper's mobility index[4]. Any of these data sets can be used as benchmark mobility data set. Although, these data resources were created in 2020 in response to COVID19 pandemic, their data are continuing to be generated beyond the pandemic. We select Google's CCMR as it is compatible with venue check-ins data and generated from the same type of data used in Google Maps, the most popular navigation application in the world (Williams 2018).

Mobility data on *Google*'s CCMR are collected from users who have enabled location history to their accounts. Data are first anonymized, then aggregated to illustrate general changes in resident mobility in specific geographic regions (e.g., city, state, country). Data show changes in visits to six major venue groups (Grocery & pharmacy, Parks, Transit stations, Retail & recreation, Residential, and workplaces). Because *Google* CCMR and *Foursquare* grouping of major venue categories are not identical, we group *Foursquare*'s sub-venue types following the corresponding *Google* CCMR major categories. For some of *Google*'s major venue groups (e.g., Residential, Transit stations, Workplaces), this process is straightforward (Residence, Travel & Transport, and Professional & Other Places, respectively). Sub-venue types in other major *Foursquare* categories are then examined and regrouped, according to *Google*'s major venue groups (Grocery & pharmacy, Parks, Transit stations, Retail & recreation). For instance, venues belonging to retail, recreation, and eateries are grouped under Retail & recreation. The major category Parks includes official national parks and not the general outdoors found in rural areas. Table 1 shows the major venue groups of *Google*'s CCMR and examples of the mapped sub-venue types from *Foursquare*.

To reflect changes in visits to major venue categories, check-in records are aggregated in the same way as *Google*'s CCMR. We then compare *Google* CCMR and *Foursquare* venue check-in data by visually inspecting data plotted on the line charts. Correlation analysis was used to assess the similarity in these two data sets (Jiang 2018). If *Foursquare* venue check-ins produce similar patterns to *Google*'s CCMR, then we deduce that they can effectively capture resident movement.

---

[2] https://www.google.com/covid19/mobility.

[3] https://covid19.apple.com/mobility.

[4] https://citymapper.com.

Table 1 Category mapping rules of check-ins

Google Category	Sample *Foursquare* sub-venue type
Grocery & Pharmacy	Supermarket, Fruit & Vegetable Store, Grocery Store, Liquor Store, Drugstore, Pharmacy
Parks	Botanical Garden, National Park, Nature Preserve, Park
Retail & Recreation	Clothing Store, Department Store, Miscellaneous Shop, Shopping Mall, Scenic Lookout, Waterfront, Recreation Centre, Lake, Garden, Beach, Campground, Restaurant, Cafe
Residential	Housing Development, Home, Residential Building
Transit stations	Bus Stop, Metro Station, Tram Station, Airport, Pier, Taxi Stand
Workplace	Government Building, Office, Business Center, Warehouse, Industrial Estate, Community Center, Convention Centre, Factory

3.3 Exploratory analysis

We conduct series of analyses using spatial (GPS), temporal (date, time), and contextual (venue type) check-in information to demonstrate the applicability of venue check-in data in domestic tourism management. We use probabilistic topic modeling based on latent Dirichlet allocation (LDA) (Vu et al. 2019) to explore the major daily resident activity themes in a specific geographical area (city). LDA is a powerful topic modeling technique, which is capable of discovering underlying topics in a large collection of documents. LDA has been widely applied to textual data in tourism literature (Calheiros et al. 2017; Luo et al. 2020).

A recent study has shown that LDA could also be applied to check-in data to reveal underlying activity preferences of tourists (Vu et al. 2019), where with each venue type treated as a word, and user daily check-in records treated as a document. Occurrences of venue types for each day are examined to identify major activity themes in each document and the entire data set, which enable multiple activities to be considered concurrently to reveal dominant activity themes, rather than separately as individual visits to specific venues as in *Google*'s CCMR (Aktay et al. 2020). Further insights can be revealed by analysis of check-in records and visualization using GIS, such as Google Earth (https://www.google.com/earth/).

3.4 Case study

We first present our experimental data sets involving venue check-ins generated by residents of several representative countries. Comparative analysis is then carried out to assess the reliability of venue check-ins in capturing residents' mobility. The potential of venue check-in data in providing insights to support domestic tourism development is then discussed through exploratory analysis of resident's travel.

Table 2 Statistics of the collected check-in data sets in the studied countries

Country	No. of Check-ins	No. of Venues	Users	Male (%)	Female (%)
Singapore	13,292	3535	145	56.6	43.4
UK	24,043	8123	612	77.4	22.5
USA	257,560	92,781	3998	67.4	32.6
Total	294,895	104,439	4755	66.9	33.1

Table 3 Correlation analysis of venue check-ins and *Google*'s CCMR (*p*-value in parentheses). Abbreviations: Ret. & Rec (Retail & recreation), Gro. & Pha. (Grocery & Pharmacy), UK (United Kingdom), USA (United States of America)

Country	Gro. & Pha.	Parks	Ret. & Rec.	Residential	Stations	Workplace
Singapore	0.07 (0.26)	0.11 (0.07)	0.58 (<0.01)*	0.04 (0.50)	0.10 (0.11)	0.35 (<0.01)*
UK	0.14 (0.03)*	0.07 (0.26)	0.88 (<0.01)*	−0.21 (<0.01)*	0.75 (<0.01)*	0.65 (<0.01)*
USA	0.51 (<0.01)*	0.24 (<0.01)*	0.84 (<0.01)*	−0.09 (0.13)	0.89 (0.00)*	0.76 (<0.01)*

*Significant at $p \leq 0.05$

3.5 Data collection

Venue check-in tweets were collected using *Twitter's* API for 6-month period spanning May to October 2020, with no bounding box, to identify *Twitter* users who also used *Foursquare* check-in features. Because the deployment period is in the middle of the pandemic when many international borders were closed, we deduce that users who generated check-ins within a country were local residents. We focus analysis on Tweets from the USA, UK, and Singapore, where *Foursquare* is a popular venue check-in platform (Humeres 2014). Using the *getUserTimeline* function we then extract all check-in tweets from *Twitter* timelines of these users, and collect check-in meta-data (e.g., time, location) and venue (e.g., venue category), to capture all historical check-in tweets beyond our streaming period.

Since, *Google*'s CCMR is a new platform, whose data are available from early 2020 onward. We retain check-in data from January to October 2020 to compare with our benchmark data set, *Google*'s CCMR. In total 294,895 check-ins were generated for 104,439 venues, from 4755 residents (Table 2). There are more USA check-ins and users than for the UK, for which there are more data returns than for Singapore. *Google's* CCMR for these countries were also retrieved. We developed Matlab and Python scripts for data collection, processing, and analysis.
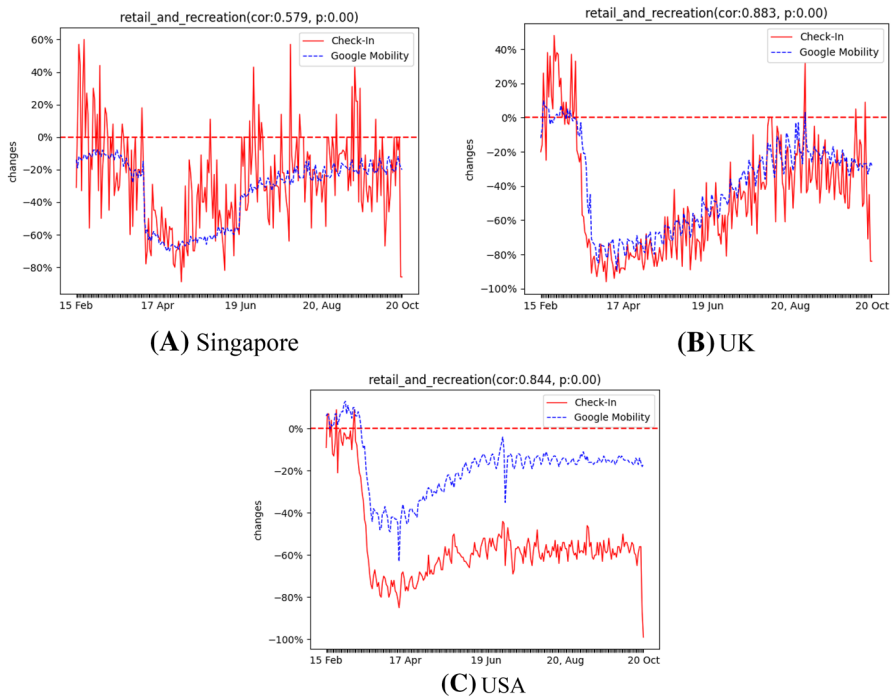
Fig. 1 Venue check-ins and Google's CCMR with strong correlations

3.6 Mobility assessment

Daily changes in visits to venues of major categories are computed as outlined in 'Benchmark assessment data sets.' Correlation analysis ($p \leq 0.05$) were performed for time-series data from February 15, 2020 (when *Google's* CCMR data were first available) to October 20, 2020 (when check-in collection was completed). Many strong positive correlations occur, between check-in data and *Google's* CCMR for major categories associated with resident movement, such as Retail & recreation, Stations, and Workplace, especially for the UK and USA (Table 3). In some cases, the *p* value was significant, but the correlation was low (e.g., UK for Grocery/Pharmacy).

A plot of check-in data against *Google's* CCMR for cases with a strong correlation, such as Retail & recreation (Fig. 1), and low or no correlation, such as Residential (Fig. 2) depict decreases in resident visits to retail & recreation venues in each country, with the largest drop in March and April. Although the magnitudes of change vary slightly, clear correlations are apparent between the two data sets. High positive correlations also occur for some categories, such as Stations and Workplace. Data size appears to influence the ability to capture resident mobility, with the USA tending to have higher correlation values than the UK or Singapore, although clear correlations exist for some categories in Singapore (e.g., Retail & recreation, Workplace). Minimal or no clear correlations are apparent between data
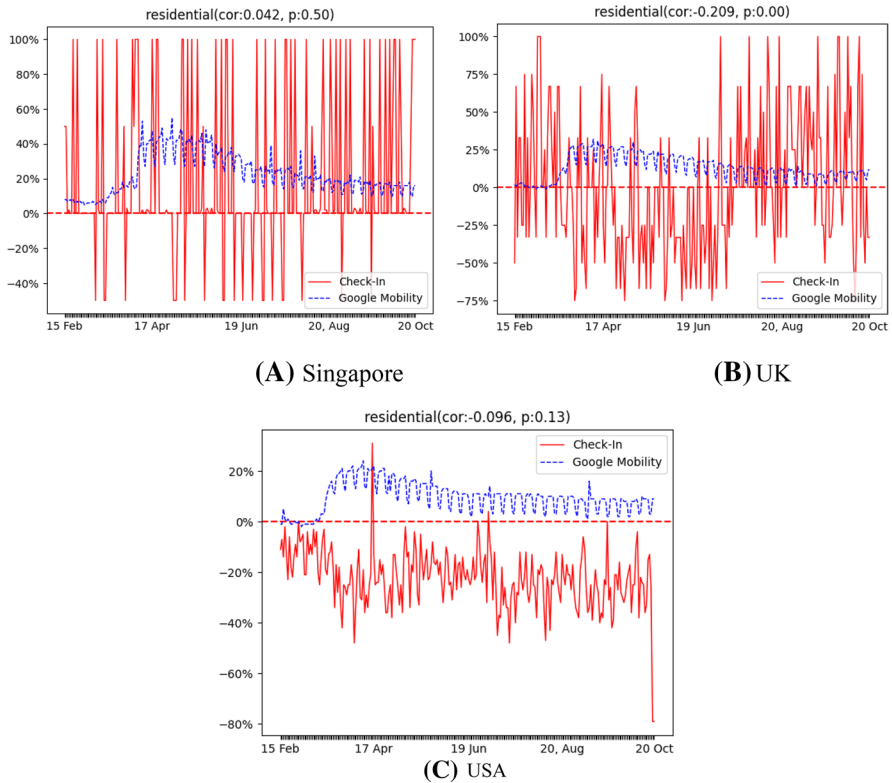
(A) Singapore **(B)** UK

(C) USA

Fig. 2 Venue check-ins and Google's CCMR with low or no correlations

sets for Parks and Residential, probably because of the small number of venues in these categories (resulting in limited check-in data). Negative correlations are apparent between data sources in the Residential category (Fig. 2), with CCMR indicating increased data generated for this category. While this suggests that residents tended to remain home, it might also be because residents were less likely to use the check-in application *Foursquare*. Motivations for residents to check-in include establishing (and maintaining) an image of themselves, so sharing a 'home status' might be less attractive to *Foursquare* users

3.7 Daily activity analysis

To demonstrate how check-in data can provide insights into resident behavior, we analyze the UK check-in data generated from March to October 2020 (n = 16,155) for 10 major *Foursquare* categories. Check-in data were converted into a suitable representation for topic modeling with LDA in a manner similar to that of Vu et al. (2019). After removing documents with less than 5 check-in records from the modeling process, approximately 2000 daily check-in records remained. Infrequently visited venues (those in less than 1% of all documents) were also removed, leaving
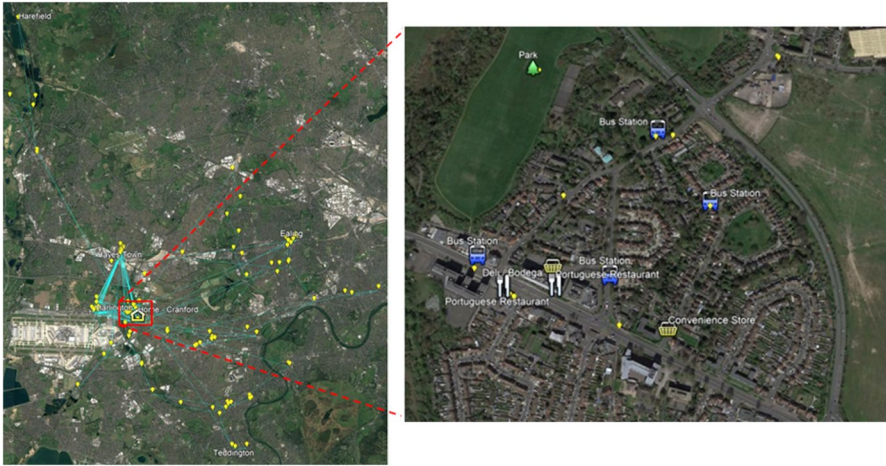
(A) Activity distribution

(B) Word cloud of topic

Fig. 3 Representative daily activities of UK residents

150 venue types. We used a tenfold cross-validation technique to evaluate the LDA model (Han et al. 2012) with varied topic numbers using a performance measure called perplexity (a low perplexity value indicates high model performance); we identified 12 topics to achieve relatively good performance.

Fig. 4 Daily activity of a resident

Figure 3a shows the heatmap probability distribution of major venue categories with respect to the 12 topics (T1 to T12). Higher probabilities depicted by progressively dark-colored cells. Some topics are strongly associated with visiting a specific venue category, such as topic T7 is strongly associated with Travel & Transport venues or T12 is strongly associated with Professional & Other Places venues. Other topics are highly associated with multiple venue categories, such as Shop & Services and Outdoors & Recreation (Topics T1 and T2).

In general, high probability values indicate that the daily activities of UK residents center on Outdoor & Recreation and Shop & Service venues. Few residents visited College & University, Event, and Nightlife Spot during this time. High probability Travel & Transport (e.g., T5, T7, T9) topics indicate the ongoing importance of daily travel for some residents—one would hope because they were essential workers or traveled for business, rather than flagrantly disregarding a government directive of no travel. Cross-referencing this with Fig. 3b, resident travel activities mainly involve buses (T7), trains (T9), and roads (T5). Topic T12 captures the activities of residents who work in Factory and Office environment settings. Pubs and bars remained important components for some UK residents (high probability in T3).

Travel behavior and activity in an area can be revealed by examining user check-in records. For demonstration purposes, we analyze the check-in records of one resident in Cranford, near London's Heathrow Airport (Fig. 4 left) who made 1015 check-ins between March and October 2020. This resident mainly traveled close to home and nearby suburbs (e.g., Harlington, Hayes), but occasionally further afield (e.g., Ealing, Harefield, Teddington), commuting primarily by bus (as indicated by many recorded check-ins at bus stations)—those close to home are shown in Fig. 4 right. Note that the exact home location of the resident is not shown due to privacy issue. The resident also frequently visited grocery stores and Portuguese restaurants near home.
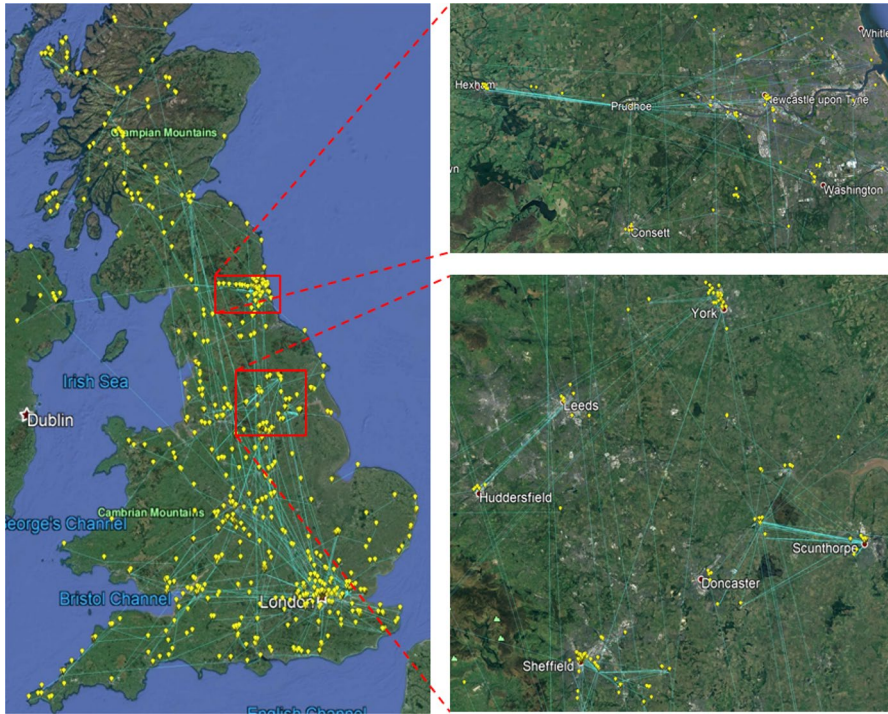
Fig. 5 Long-distance resident movements

3.8 Domestic travel behavior

To demonstrate how resident check-in data can capture domestic daily travel and tourism activities of or exceeding 50 km, we use UK check-in data from March to October 2020, visualizing travel trajectories using Google Earth (Fig. 5a–c). Figure 5 left depicts travel trajectory at a country scale to provide an overview of how residents move between major cities. Because travel between distant cities within a day, such as between London and northern cities, is typically impractical by road, air travel is required. Figure 5 top right and bottom right show travel trajectories between cities: routes between Hexbam and Prudhoe (Fig. 5 top right), and Huddersfield and Leeds (Fig. 5 bottom right) are busy with many residents who travel daily.

Residents may travel for purposes other than leisure, such as for business or visiting family. To differentiate tourism-related travel from these other activities we regard long-distance travel from home which involves visits to parks and recreational venues, or stays in a hotel, as likely tourism related. Among the UK residents in our collected data set (Table 2), at least 71 UK users (approximately 12%) checked-in at least once in a hotel.

Figure 6 shows the travel itinerary of a London tourist who visited Newquay. Nine (1–9 in chronological order) check-ins were made on October 13, 2020: (1)

Fig. 6 1-day itinerary of a tourist

into a hotel in Newquay, (2) a grocery then (3) liquor store, (4) an Italian restaurant, (5) a brewery (Fig. 6 bottom right), the nearby St. Ives (Fig. 6 bottom left), visiting two pubs (6, 7), and 8) St. Ives harbor before returning to Newquay, before 9) finally checking in at a cafe before finishing the day (Fig. 6 bottom right).

3.9 Discussion and implications

The correlation analysis with benchmark data sets (see 'Mobility assessment') reveals how check-in data can be meaningfully used to capture and identify changes in resident mobility. However, the ability of these data to achieve this varies with venue type. Check-in data are more able to capture visits to Retail & Recreation, Stations, and Workplace categories, covering a wide range of venues related to resident daily activities, than those related to Grocery & Pharmacy, Parks, and Residential categories. Several factors affect data set size, including the survey country or region, number of users at that location, and the focus of data collection. Countries with larger populations and more users such as the USA will obviously have larger data sets than those which do not. Nevertheless, data for Singapore still reveal important insights into resident mobility, with low but statistically significant correlations found for certain major categories. We recommend that check-in data can be utilized to complement other data sources to support the decision-making processes of tourism managers, where *Foursquare* check-ins are popular among residents, such as the USA, UK, Singapore, Japan, Malaysia, and Thailand (Vu et al. 2019, 2020).

It must be borne in mind, however, that mobility patterns generated from these data may not be representative of the entire population: *Foursquare* users are typically young. Besides, studies should first assess whether *Foursquare* check-in data, or those from some other platform, are suitable for a specific country before analysis—for instance, *Foursquare* is not popular in China, but *Dianping* is. The reliability of these platforms in capturing mobility and activities of people should also be validated, as presented in this paper, before adopting them for practical applications.

Besides the mobility patterns, tourism managers may be more interested in identifying in additional insights into residents' activities. Using topic-modeling (see 'Daily activity analysis'), we identify major activity themes for UK residents, with those related to Outdoor & Recreation remaining essential to daily life; residents liked to visit venues such as beaches (T1), harbors/marinas (T4), and parks, trails, and bridges (T8, T11). Shop & Service venues such as supermarkets (T2), shopping malls (T6), and stores (T10) were also often visited by residents. Travel & Transport, especially by train (T9) or bus (T7), is an important component of daily life for residents who commute for essential purposes. If local governments study resident movement and activity within their jurisdiction, they can tailor support to improve the livelihoods of residents and well as promoting domestic tourism activities. Long-distance movement (Fig. 5, see 'Domestic travel behavior') reveals how check-in data can capture resident travel flows in the same way as other data sources (Chua et al. 2016; Li et al. 2020; Raun et al. 2016; Versichele et al. 2014; Vu et al. 2015). Check-in data also provides insights into tourist activities at visited locations because of venue information (Fig. 6). Distinguishing tourism-related travel from daily life activities is also possible with user home location information (check-in record) or city of origin (*Foursquare* profile), which cannot easily be done using other data sources.

There are several limitations with this paper as follows. We collected data using *Twitter*'s API, the quota limit of which returns a limited portion of available data generated in real time. If more data are needed (e.g., such as for small countries such as Singapore) then researchers could consider premium access to *Twitter*'s API (Twitter 2020) to provide all available data. Because of length limitations and our research focus (assessing the data source reliability), and the fact that this is essentially a proof-of-concept study, we did not perform in-depth analysis of tourist activities. Such insights could be obtained using existing travel analysis and modeling techniques (Chua et al. 2016; Li et al. 2020; Raun et al. 2016; Versichele et al. 2014; Vu et al. 2015). Future work focusing on studying the domestic behavior of tourists could also identify travel-related check-ins before conducting in-depth analysis for specific domestic tourism applications.

4 Conclusions

Mobility and movement have always been central to travel. Disruptions to travel mobility are not new, but a recurring certainty (Adey et al. 2021). For tourism, examples of disruptions include health risks, climate change, natural disasters, crime, political instability, and terrorism (Karl et al. 2020). During these periods,

international travel tends to be limited, and domestic tourism plays an important role in economic development of many countries. Understanding the mobility and activities of residents is crucial for tourism managers to manage and develop domestic tourism during disruption and beyond. VR-SMD in the form of venue check-ins represent a relatively new type of data that can overcome the limitations of most traditional approaches to capturing resident mobility (e.g., mobile phone tracking, micro-blogging data), which lack contextual information on resident activities. Until now, the ability of VR-SMD data to describe resident movement for domestic tourism development has not been realized. In this paper, we presented an approach to assessing the reliability of venue check-in data in capturing resident mobility and demonstrated how these data can provide valuable insights into residents' daily and tourism activities.

A natural extension of this work would be to assess the ability of venue check-in data in other countries and regions to provide similar insights into the movement patterns of residents. An in-depth analysis of domestic tourism activities could be conducted, where existing travel analysis and modeling techniques (Chua et al. 2016; Li et al. 2020; Raun et al. 2016; Versichele et al. 2014; Vu et al. 2015) could be employed for practical implementations (Vu et al. 2019, 2020). Changes in residents' mobility between normal and disruption periods can also be studied, so that domestic tourism management strategy can be appropriately adjusted to maximize their economic benefits.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

ACT (2021) Check in CBR App. https://www.covid19.act.gov.au/business-and-work/check-in-cbr
Adey P, Hannam K, Sheller M, Tyfield D (2021) Pandemic (im) mobilities. Mobilities 16(1):1–9. https://doi.org/10.1080/17450101.2021.1872871

Aktay A, Bavadekar S, Cossoul G, Davis J, Desfontaines D, Fabrikant A, Gabrilovich E, Gadepalli K, Gipson B, Guevara M, Kamath C et al (2020) Google COVID-19 community mobility reports: anonymization process description (version 1.0).https://arxiv.org/abs/2009.01265

Assaf AG, Kock F, Tsionas M (2022) Tourism during and after COVID-19: an expert-informed agenda for future research. J Travel Res 61(2):454–457. https://doi.org/10.1177/00472875211017237

Baggio R, Scaglione M (2018) Strategic visitor flows and destination management organization. Inform Technol Tourism 18(1):29–42. https://doi.org/10.1007/s40558-017-0096-1

Barchiesi D, Moat HS, Alis C, Bishop S, Preis T (2015) Quantifying international travel flows using Flickr. PLOS One 10(7):e0128470. https://doi.org/10.1371/journal.pone.0128470

Calheiros AC, Moro S, Rita P (2017) Sentiment classification of consumer-generated online reviews using topic modeling. J Hosp Mark Manag 26(7):675–693. https://doi.org/10.1080/19368623.2017.1310075

Chua A, Servillo L, Marcheggiani E, Moere AV (2016) Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy. Tour Manag 57:295–310. https://doi.org/10.1016/j.tourman.2016.06.013

Cranshaw J, Lindqvist J, Wiese J, Hong J, Zimmerman J (2011) I'm the Mayor of my House: examining why people use Foursquare—A Social-driven location sharing application. Proceedings of the SIGCHI conference on human factors in computing systems 2409–2418.Association for Computing MachineryUSAhttps://doi.org/10.1145/1978942.1979295

Demmler GJ, Ligon BL (2003) Severe acute respiratory syndrome (SARS): a review of the history, epidemiology, prevention, and concerns for the future. Semin Pediatr Infect Dis 14(3):240–244. https://doi.org/10.1016/S1045-1870(03)00056-6

Foursquare (2020) Venue Categories. https://developer.foursquare.com/docs/build-with-foursquare/categories/

Ghermandi A, Camacho-Valdez V, Trejo-Espinosa H (2020) Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tour Manag 77:104002. https://doi.org/10.1016/j.tourman.2019.104002

Grantz KH, Meredith HR, Cummings DA et al (2020) The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. Nat Commun 11(1):4961. https://doi.org/10.1038/s41467-020-18190-5

Han J, Kamber M, Pei J (2012) Data Mining: concepts and techniques, 3rd edn. Morgan Kaufmann, Waltham, MA, p 02451

Huang X, Li Z, Jiang Y, Li X, Porter D (2020) Twitter reveals human mobility Dynamics during the COVID-19 pandemic. PLOS One 15(11):e0241957. https://doi.org/10.1371/journal.pone.0241957

Humeres MFJ (2014) Power centrality as a relational measure of urban hierarchy: testing the splintering urbanism theory with social media data from Santiago de Chile. PhD thesis. M.C.P., Massachusetts institute of technology, department of urban studies and planning. https://dspace.mit.edu/handle/1721.1/87523

Jiang W (2018) A correlation coefficient for belief functions. Int J Approx Reason 103:94–106. https://doi.org/10.1016/j.ijar.2018.09.001

Jin L, Long X, Zhang K, Lin YR, Joshi J (2016) Characterizing users' check-in activities using their scores in a location-based social network. Multimed Syst 22(1):87–98. https://doi.org/10.1007/s00530-014-0395-8

Joo HH, Kang HG, Moon JJ (2014) The effect of rain on the decision to visit a theme park. Asia Pac J Tour Res 19(1):61–85. https://doi.org/10.1080/10941665.2012.724020

Karl M, Muskat B, Ritchie BW (2020) Which travel risks are more salient for destination choice? an examination of the tourist's decision-making process. J Destin Mark Manag 18:100487. https://doi.org/10.1016/j.jdmm.2020.100487

Li Y, Zeng Y, Liu G, Lu D, Yang H et al (2020) Public awareness, emotional reactions and human mobility in response to the COVID-19 outbreak in China—a population-based ecological study. Psychol Med 52(9):1793–1800. https://doi.org/10.1017/S003329172000375X

Luo JM, Vu HQ, Li G, Law R (2019) Tourist behavior analysis in gaming destinations based on venue check-in data. J Travel Tour Mark 36(1):107–118. https://doi.org/10.1080/10548408.2018.1494088

Luo JM, Vu HQ, Li G, Law R (2020) Topic modelling for theme park online reviews: analysis of Disneyland. J Travel Tour Mark 37(2):272–285. https://doi.org/10.1080/10548408.2020.1740138

Mou N, Yuan R, Yang T, Zhang H, Tang JJ, Makkonen T (2020) Exploring spatio-temporal changes of city inbound tourism flow: the case of Shanghai, China. Tour Manag 76:103955. https://doi.org/10.1016/j.tourman.2019.103955

NHS (2020) Venue check-in for organisations-NHS COVID-19 app support-nhs.uk. https://covid19.nhs.uk/venue-check-in-businesses.html https://doi.org/10.1080/13683500.2022.2081789

Orellana D, Bregt AK, Ligtenberg A, Wachowicz M (2012) Exploring visitor movement patterns in natural recreational areas. Tour Manag 33(3):672–682. https://doi.org/10.1016/j.tourman.2011.07.010

Pandey DK, Kumar R (2022) Russia-Ukraine War and the global tourism sector: a 13-day tale. Curr Issues Tour 26(5):692–700. https://doi.org/10.1080/13683500.2022.2081789

Patil S, Norcie G, Kapadia A, Lee A (2012) "Check out where i am!" location-sharing motivations, preferences, and practices. In: CHI'12 extended abstracts on human factors in computing systems. Association for Computing Machinery, USA https://doi.org/10.1145/2212776.2223742

Payntar ND, Hsiao WL, Covey RA, Grauman K (2021) Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. Tour Manag 82:104165. https://doi.org/10.1016/j.tourman.2020.104165

Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. Tour Manag 57:202–212. https://doi.org/10.1016/j.tourman.2016.06.006

Salas-Olmedo MH, Moya-Gómez B, García-Palomares JC, Gutiérrez J (2018) Tourists' digital footprint in cities: comparing big data sources. Tour Manag 66:13–25. https://doi.org/10.1016/j.tourman.2017.11.001

Shoval N, Ahas R (2016) The use of tracking technologies in tourism research: the first decade. Tour Geogr 18(5):587–606. https://doi.org/10.1080/14616688.2016.1214977

Shoval N, McKercher B, Ng E, Birenboim A (2011) Hotel location and tourist activity in cities. Ann Tour Res 38(4):1594–1612. https://doi.org/10.1016/j.annals.2011.02.007

Sobolevsky S, Bojic I, Belyi A, Sitko I, Hawelka B, Arias JM, Ratti C (2015) Scaling of City Attractiveness for Foreign Visitors Through Big Data of Human Economical and Social Media Activity. In: 2015 IEEE International Congress on Big Data, 2015, pp. 600–607. https://doi.org/10.1109/BigDataCongress.2015.92

Statista (2020) Travel & tourism: global economic impact 2006–2019 Statista. https://www.statista.com/statistics/233223/travel-and-tourism--total-economic-contribution-worldwide. Accessed 02 Mar 2021

Stienmetz JL, Fesenmaier DR (2019) Destination value systems: modeling Visitor Flow structure and economic impact. J Travel Res 58(8):1249–1261. https://doi.org/10.1177/0047287518815985

TRA (2020) Moving forward: the role of domestic travel in Australia's tourism recovery. https://www.tra.gov.au/ArticleDocuments/185/Moving-Forward-2020.pdf.aspx. Accessed 4 Feb 2021

Twitter (2020) Introducing twitter premium APIs. https://developer.twitter.com/en/products/twitter-api/premium-apis (Accessed on February 4, 2021)

Versichele M, De Groote L, Bouuaert MC, Neutens T, Moerman I, Van de Weghe N (2014) Pattern mining in tourist attraction visits through association rule learning of bluetooth tracking data: a case study of Ghent, Belgium. Tour Manag 44:67–81. https://doi.org/10.1016/j.tourman.2014.02.009

Vu HQ, Li G, Law R, Ye BH (2015) Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. Tour Manag 46:222–232. https://doi.org/10.1016/j.tourman.2014.07.003

Vu HQ, Li G, Law R, Zhang Y (2018) Tourist activity analysis by leveraging mobile social media data. J Travel Res 57(7):883–898. https://doi.org/10.1177/0047287517722232

Vu HQ, Li G, Law R (2019) Discovering implicit activity preferences in travel itineraries by topic modeling. Tour Manag 75:435–446. https://doi.org/10.1016/j.tourman.2019.06.011

Vu HQ, Li G, Law R (2020) Cross-country analysis of tourist activities based on venue-referenced social media data. J Travel Res 59(1):90–106. https://doi.org/10.1177/0047287518820194

Wellenius GA, Vispute S, Espinosa V, Fabrikant A, Tsai TC et al (2020) Impacts of state-level policies on social distancing in the United States using aggregated mobility data during the COVID-19 pandemic. Nat Commun 12:3118 (**arXiv preprint arXiv:2004.10172**)

Williams R (2018) Google Maps Rated as No. 1 Navigation App, Survey Says. https://www.marketingdive.com/news/google-maps-rated-as-no-1-navigation-app-survey-says/527525/ (accessed February 4, 2021)

Wörndl W, Hefele A, Herzog D (2017) Recommending a sequence of interesting places for tourist trips. Inform Technol Tourism 17(1):31–54. https://doi.org/10.1016/j.ejor.2014.04.006

WTTC (2020) Travel & tourism economic impact - World Travel & Tourism Council (WTTC). https://wttc.org/Research/Economic-Impact. Accessed 02 Mar 2021

Xiong C, Hu S, Yang M, Luo W, Zhang L (2020) Mobile Device Data Reveal the Dynamics in a Positive Relationship Between Human Mobility and COVID-19 Infections. Proceedings of the National Academy of Sciences 117(44):27087–27089. https://doi.org/10.1073/pnas.2010836117

Zheng YT, Zha ZJ, Chua TS (2011) Research and applications on georeferenced multimedia: a survey. Mul-
      timed Tools Appl 51(1):77–98. https://doi.org/10.1007/s11042-010-0630-z
Zheng W, Huang X, Li Y (2017) Understanding the tourist mobility using GPS: where is the next place?
      Tour Manag 59:267–280. https://doi.org/10.1016/j.tourman.2016.08.009
Zhou Y, Xu R, Hu D, Yue Y, Li Q, Xia J (2020) Effects of human mobility restrictions on the spread of
      COVID-19 in Shenzhen, China: a modelling study using mobile phone data. Lancet Digit Health
      2(8):e417–e424. https://doi.org/10.1016/S2589-7500(20)30165-5

Authors and Affiliations

Huy Quan Vu[1] · **Shah Jahan Miah[2]** · **Haiyang Xia[3]** · **Gang Li[4]** ·
Birgit Muskat[3] · **Rob Law[5]**

✉ Huy Quan Vu
    q.vu@deakin.edu.au

    Shah Jahan Miah
    shah.miah@newcastle.edu.au

    Haiyang Xia
    haiyang.xia@anu.edu.au

    Gang Li
    gang.li@deakin.edu.au

    Birgit Muskat
    birgit.muskat@anu.edu.au

    Rob Law
    roblaw@um.edu.mo

[1] Department of Information Systems and Business Analytics, Deakin University, Melbourne,
     Australia

[2] Newcastle Business School, University of Newcastle, Newcastle, Australia

[3] Research School of Management, Australian National University, Canberra, Australia

[4] School of Information Technology, Deakin University, Melbourne, Australia

[5] Asia-Pacific Academy of Economics and Management, Department of Integrated Resort
     and Tourism Management, Faculty of Business Administration, University of Macau, Macau,
     China