Received: 25 May 2022 | Revised: 31 August 2022 | Accepted: 16 September 2022
DOI: 10.1111/gec3.12663

REVIEW ARTICLE

WILEY

Modeling activity spaces using big geo-data: Progress and challenges

Yihong Yuan[1] | Yang Xu[2]

[1]Department of Geography and Environmental Studies, Texas State University, San Marcos, Texas, USA

[2]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Correspondence
Yihong Yuan, Department of Geography and Environmental Studies, Texas State University, San Marcos, TX, USA.
Email: yuan@txstate.edu

Abstract

The growing availability of big geo-data, such as mobile phone data and location-based social media (LBSM), provides new opportunities and challenges for modeling human activity spaces in the big data era. These datasets often cover a large sample size and can be used to model activity spaces more efficiently than traditional travel surveys. However, these data also have inherent limitations, such as the lack of reliable demographic information of individuals and a low sampling rate. This paper first reviews the strengths and weaknesses of various internal and external activity space indicators. We then discuss the pros and cons of using various new data sources (e.g., georeferenced mobile phone data and LBSM data) for activity space modeling. We believe this review paper is a valuable reference not only for researchers who are interested in activity space modeling based on big geo-data, but also for planners and policy makers who are looking to incorporate new data sources into their future workflow.

KEYWORDS
activity space modeling, big geo-data, location-based social media, mobile phone data

1 | INTRODUCTION

In geography and transportation studies, modeling activity spaces is a crucial topic when studying the characteristics of individual behaviors and collective urban mobilities (Golledge & Stimson, 1997; Jones et al., 1990; Silm & Ahas, 2014; Yuan & Raubal, 2016). For example, by understanding urban residents' activity space patterns, urban planners can evaluate residents' accessibility to critical resources, such as healthcare facilities, public transportation, and leisure facilities, based on the availability of resources within an individual's activity space (Raskind et al., 2020; Sherman et al., 2005a, 2005b). Previous studies defined activity space as the local areas within which people travel during their daily activities (Mazey, 1981). Related concepts include, but are not limited to, *space-time prisms* (Hägerstrand, 1970), the *awareness space* (Brown & Moore, 1970), and the *action space* (Horton & Reynolds, 1971). A large branch of activity space studies have focused on approximating the external morphology (e.g., size and shape) and the internal structure (e.g., the regularity of visitation patterns) of activity spaces (Schönfelder & Axhausen, 2002; Sherman et al., 2005b).

With the development of mobile positioning technologies (e.g., smart phones and built-in GPS devices), there is a growing body of research applying data generated by these new technologies to quantitively model the morphology and structure of activity spaces, as well as investigate how activity spaces form in different urban environments. Example datasets include, but are not limited to, cell phone call detailed records (CDRs) (Ahas, 2005; Ahas et al., 2015), mobile signaling data (MSD) (Xu, Shaw, Zhao, et al., 2016), location-based social media (LBSM) check-in data (Yuan & Wang, 2018), and smart card transaction data (Zhang et al., 2021). Compared to traditional surveys and travel diaries, these datasets provide a valuable resource for understanding human activity spaces on a large spatio-temporal scale, but there are also data quality and uncertainty issues associated with big geo-data (Wesolowski et al., 2013; Yuan et al., 2018). It is crucial to understand the impact of these issues when applying such data to human activity space studies.

The following sections of this paper are organized as follows. Section 2 introduces commonly used activity space metrics and indicators. Section 3 discusses how various types of big geo-data, with a focus on mobile phone data and LBSM data, have been used for activity space modeling. Section 4 concludes this paper and summarizes the contributions.

2 | ACTIVITY SPACE INDICATORS

2.1 | External indicators

Over the years, numerous internal and external indicators have been introduced to characterize the geometric properties of an individual's daily activities. Here we define external indicators as the metrics measuring the external morphology, such as the size, shape, and orientation, of human activity space. We define internal indicators as those that analyze the structure and formation of activities, such as the randomness and regularity of activity spaces.

A widely used metric in activity space research is the standard deviational ellipse (SDE). This is because an ellipse can effectively capture the size, shape, and orientation of activity spaces simultaneously. Taking a collection of point events as input, the SDE seeks to define an ellipse, of which its major and minor axes tend to capture the directions along which the points have the maximum and minimum spatial dispersion (Figure 1). The center of the SDE is determined by computing the geometric center (i.e., mean center) of the points. The lengths of the major and minor axes of the SDE are determined by measuring the standard deviation of the points' coordinates along these two axes (Yuill, 1971). The eccentricity e demonstrates the elongation of the SDE. In the context of activity space research, the locations extracted from an individual's activities, as well as the frequencies of visiting these locations, are used together to determine the SDE. Thus, the SDE can not only reflect the spatial extent of a person's activities, but

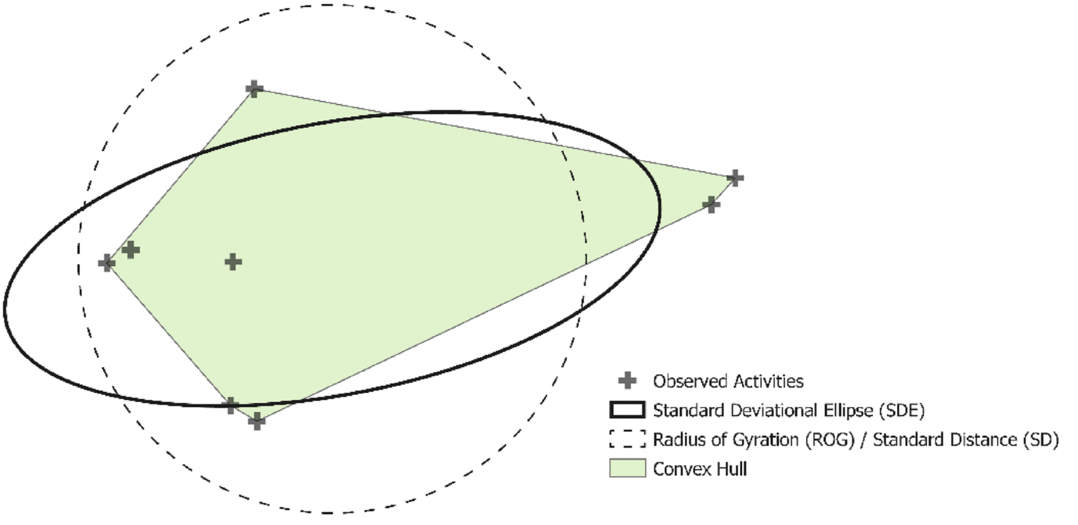

FIGURE 1 Graphical illustration of the standard deviational ellipse (SDE), radius of gyration (ROG), standard distance (SD), and Convex Hull

also the directional distribution of these activities (Gesler & Meade, 1988; Schönfelder & Axhausen, 2003; Sherman et al., 2005b; Zenk et al., 2011).

A closely related indicator is the standard distance (SD) (Figure 1). Given a collection of points, this indicator is computed as the standard deviation of the distance of the points from the mean center (i.e., the centroid of all points) (Bachi, 1963). Compared to SDE, which provides an elliptical view of human activity space, the standard distance can be considered a circular view of the spatial dispersion of one's activities (Buliung & Kanaroglou, 2006; Schönfelder & Axhausen, 2002). The standard distance is also closely related to the radius of gyration (ROG), a metric widely used in physics and complexity science. ROG has an equivalent form of standard distance when it is used to measure the spread of geographic locations and the spatial dispersion of individual travel behavior (Gonzalez et al., 2008; Xu et al., 2018). As mentioned in Schönfelder and Axhausen (2003), the mean center of an individual's activities is considered the central location of an activity space when SDE or the standard distance is applied.

In addition to the above measures, the minimum convex polygon, or convex hull (Figure 1), has also been applied to quantify an individual's activity range (Buliung & Kanaroglou, 2006; Thériault et al., 1999). This metric captures the minimum spatial area that can cover the entire set of locations where individuals practice their activities. Convex hulls are very simple to compute but are highly sensitive to outliers: one point far from the mean center can possibly cause a substantial change in this indicator.

2.2 | Internal indicators

As discussed in Golledge and Stimson (1997), there are three determinants of activity spaces: (1) home location; (2) regularly visited activity locations (Points of interest, POIs), such as the work location, shopping malls, etc.; and (3) travel between and around regularly visited locations (e.g., accessibility of public transportation near one's home). In traditional activity studies, locations where people regularly spend time are often used to analyze their socioeconomic status and lifestyle (Pendyala et al., 1991; Xu, Shaw, Fang, & Yin, 2016; Yamamoto & Kitamura, 1999). Unlike in travel surveys where home/work/POI locations are often explicitly asked, in studies that rely on big geo-data, researchers mostly need to estimate POI locations based on algorithms and rulesets, which inevitably introduces uncertainty into the analysis. For example, Phithakkitnukoon et al. (2010) identified stops (i.e., locations where one stayed longer than a time threshold) from mobile phone user trajectories and then defined home location as the most visited stop at night and work location as the most frequent stop between Monday–Friday, 8 AM–5 PM. Another study by Ahas et al. (2010) presented a process to identify various types of meaningful locations, such as home anchor points and work anchor points, using passive mobile phone data. Their methodology provided a valuable framework for identifying regularly visited locations, such as popular tourist destinations, in various applications.

It is worth noting that home locations are usually considered a "middle product" for extracting other internal indicators because it is often necessary to convert a point location to quantifiable metrics. For example, Hu et al. (2020) calculated the average distance between the home location and other visited locations to investigate the activity space structure of mobile phone users. In addition, home locations and other POIs can be used to improve the external indicators in Section 2.1 because these locations are considered focal points of human activities. Previous studies have modified the above measures by substituting the mean center or focal points of activity space by an individual's important activity locations (Dijst, 1999; Newsome et al., 1998; Xu et al., 2015).

In addition to extracting POIs, researchers also developed a series of probability-based indicators calculated based on the likelihood of users visiting different locations (Song et al., 2010; Wang & Yuan, 2021). One commonly used metric in this category is entropy (Equation (1)), which measures the probabilistic distribution of visiting different locations:

$$E = -\sum_{i=1}^{N} p_i \log_2 (p_i) \tag{1}$$

where p_i refers to the probability of a given user checking in at the same place i, and N stands for the total number of places where this user checked in. A higher entropy indicates a more randomly distributed visitation pattern; thus, it is more difficult to predict the future locations of a user with a higher entropy value. For example, the point pattern on the left in Figure 2 has a lower entropy value (i.e., a less dispersed pattern) than the one on the right.

Besides entropy, travel diversity is also a helpful indicator that measures the probabilistic likelihood of trips among different activity locations (Pappalardo et al., 2016; Xu, Xue, et al., 2021). Compared to entropy, travel diversity focuses more on the magnitude of interactions between location pairs (Equation (2)):

$$D = -\sum_{k \in C} p_k \log_2 (p_k) \tag{2}$$

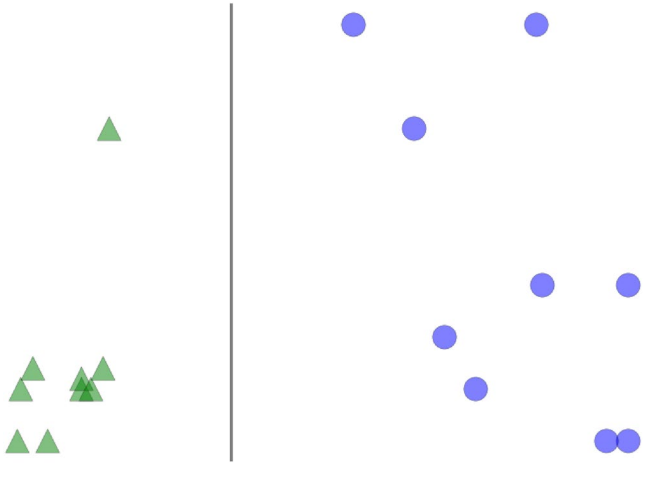

FIGURE 2 Example point patterns

where p_k is the probability of a trip happening between the kth origin-destination (OD) pair. C is a collection of OD pairs.

Another group of indicators adopts a network-based perspective to examine the connectivity and accessibility of activity spaces. Examples include road network buffer (Sherman et al., 2005b) and minimum spanning tree (Schönfelder & Axhausen, 2003). By assuming that individuals tend to travel along the shortest paths, these measures derive a collection of road networks that connect an individual's activity locations with the minimum travel cost. These measures are closely related to time-geographic measures such as potential path space and daily potential path area (Kwan, 1998, 1999).

3 | NEW DATA SOURCES FOR ACTIVITY SPACE MODELING

3.1 | Georeferenced mobile phone data and their applications

Although there remains a gap across countries, mobile phones have become a ubiquitous technology for everyday life (International Telecommunication Union, 2021). Studies have employed different types of mobile phone data in activity space research. The most widely used are call detail records (CDRs). Whenever a cellphone user engages in certain types of cellular activities (e.g., phone calls, text messaging), the time of the event and the location of the mobile phone device are recorded (Blondel et al., 2015; Xu et al., 2015). The location of the device is usually reported at the level of cellphone towers or base stations. Therefore, the spacing gap between cellphone towers—an indicator of the positional accuracy of CDRs—can vary in different areas of a city, ranging from a few hundred meters to several kilometers (Jiang et al., 2017). Mobile signaling data (MSD) is another type of mobile phone data used in existing studies (Xu, Li, et al., 2021; Yan et al., 2019). While CDRs produce sparse records for inactive users, MSD can capture user footprints in a more continuous manner through different types of signaling events, such as cellular handover, phone calls, text messaging, data connection (e.g., web browsing), and other status changes, which are triggered by telecommunication systems (Janecek et al., 2015). In other words, MSD has an improved temporal resolution over CDRs (Xu et al., 2020), especially for less active cellphone users. As with CDRs, locations in MSD are often reported at the level of cellphone towers. Some studies have introduced other types of mobile phone data, such as Erlang data (Ratti et al., 2006) and sightings data (Chen et al., 2014).

There have been numerous studies applying georeferenced mobile phone data to understand human activity spaces. We categorize them into three groups: (1) identifying the similarity and distinctions of activity space indicators among population groups; (2) understanding the spatial heterogeneity of human activity space and its connections to socioeconomic characteristics; and (3) analyzing the relation between activity spaces and social ties.

First, many studies have used georeferenced mobile phone data to calculate the activity space indicators discussed in Section 2, such as identifying important activity locations of individuals (Ahas et al., 2010; Isaacman et al., 2011), quantifying the range of activity spaces (Ahas et al., 2007; Kang et al., 2010; Phithakkitnukoon et al., 2012), and understanding how individuals allocate time across different activity locations (Bayir et al., 2009). Some of the studies specifically focused on understanding the regularity of individual human activities. By analyzing CDR data of large populations, Gonzalez et al. (2008) and Song et al. (2010) employed radius of gyration and entropy-based measures to quantify the statistical properties of individual mobility patterns. The studies found that individual human mobility exhibited a remarkable level of spatio-temporal regularity. However, some studies have also uncovered a notable level of diversity in people's use of space (Yuan et al., 2012). For example, by analyzing a CDR dataset collected in Estonia, Silm and Ahas (2014) found that ethnicity has a significant influence on the activity space of individuals. In particular, the Russian-speaking minority tended to have more confined activity spaces than the Estonian-speaking majority. The study showed that mobile phone data could be used to understand the activity space of different socio-demographic groups and their implications for social segregation and inequality.

Second, the link between people's use of space and their socioeconomic environment has led to a series of studies to analyze their connection. For example, by analyzing a CDR dataset collected in Shenzhen, China, Xu et al. (2015) introduced a modified standard distance measure to quantify the spatial dispersion of individuals' activities using home location as the reference point. The study found a north-south divide of people's activity range in Shenzhen that aligns with the socioeconomic divide of the city. Blumenstock et al. (2015) stated that behavior indicators derived from mobile phone data can be used to accurately portray the poverty and wealth level of individuals. Similar efforts were also found in several other studies (Frias-Martinez et al., 2013; Pappalardo et al., 2016), which reported varying levels of prediction accuracy and efficacy of mobility indicators. As these studies relied on machine learning models, the relationship between activity space and people's socioeconomic characteristics was not revealed explicitly. By using two CDR datasets collected in Singapore and Boston, Xu et al. (2018) derived a collection of individual activity space and mobility metrics (e.g., radius of gyration, activity entropy, travel diversity) and correlated them with the socioeconomic characteristics of individuals inferred from income and housing price data. The study found that in both cities, phone users across different socioeconomic classes exhibited a similar level of mobility diversity.

Third, as mobile phone data such as CDRs are able to capture the cellphone communications among people, the data can be used to portray activity spaces and social interactions simultaneously. Using a CDR dataset collected in Portugal, Calabrese et al. (2011) found that more than 70% of users who called each other frequently had also shared urban space at the same time. Unlike studies that examine activity space for independent individuals, this research analyzed whether the activity spaces of social contacts tended to overlap with each other. Several other studies have also employed mobile phone data to understand how people share activity spaces with others (Shi et al., 2016; Wang et al., 2015; Xu et al., 2017). These studies suggest that shared activity spaces and social ties are tightly connected, and mobile phone data have provided new opportunities for understanding this connection (Toole et al., 2015). Since mobile phone data can portray the actual and potential interactions of people in their activity spaces, the data have also been used to understand the exposure of social groups to each other and to study social segregations (Jarv et al., 2015; Leo et al., 2016; Silm & Ahas, 2014; Xu et al., 2019) (Figure 3).

3.2 | Pros and cons of using mobile phone data for modeling human activity spaces

Compared to travel surveys and questionnaires, mobile phone data offer a scalable solution for documenting the dynamics of large populations for long periods of time. The information on the when and where of individual activities are readily available, and thus can be used directly to model activity space without further digitization and geocoding. Therefore, mobile phone data have great potential to support large-scale spatial applications, such as transportation and urban planning (Chen et al., 2014), tourism analysis (Xu, Xue, et al., 2021), assessment of pollution exposure (Nyhan et al., 2016), and monitoring of human dynamics during public health crises such as COVID-19 (Huang et al., 2022). Location data can also be collected through various smart phone applications, and this data have become valuable sources for human mobility analysis because they can provide more precise location information

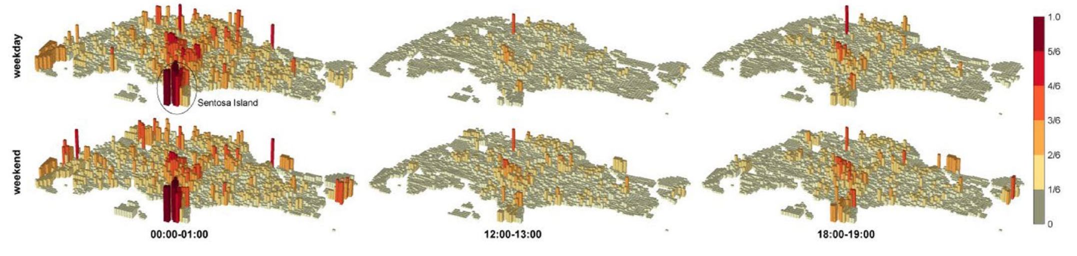

FIGURE 3 Use of mobile phone data to quantify social segregation across urban locations in Singapore. The graphic was reproduced from Figure 5(c) in Xu et al. (2019) with the permission of the authors/journal

than travel surveys can and more background detail about the participants than CDR data can (Ren et al., 2022; Xu, 2021).

There are also many challenges when mobile phone data are used to model activity space. First, mobile phone data are usually reported at the level of cellphone towers. Given the positional inaccuracy and other uncertainty issues (e.g., oscillation of cellphone signal, also known as the ping-pong effect), the location observations in the data do not always reflect the actual locations of individuals (Xu et al., 2020). Second, some of the location recordings in the data might not reflect individuals' meaningful activity locations (e.g., a pass-by location when a person initiates a call on a subway). Moreover, mobile phone data usually fall short of collecting sociodemographic characteristics of individuals, and the types of activities and trip purposes are not reported in the data. This lack of demographic information poses a considerable challenge to activity space research when activity semantics and the background of travelers are critical pieces of information. Data fusion techniques (e.g., combining CDR with census data) and novel approaches for deriving activity semantics are possible solutions to address this limitation.

3.3 | LBSM and other datasets for modeling human activity spaces

LBSM is defined as "Social Network Sites (SNS) that include location information" (Roick & Heuser, 2013). In the big data era, LBSM data are widely used for modeling human activity patterns and the perceptions of places (Sui & Goodchild, 2011; Wu et al., 2014). LBSM data usually cause fewer privacy concerns than georeferenced mobile phone data and can be obtained through application program interfaces (APIs). LBSM users can also publish more detailed background information such as their age, gender, education, and employment on their public profile (Fohringer et al., 2015; Mancosu & Bobba, 2019; Yuan et al., 2018). The geolocations in LBSM can either be coordinates acquired from built-in smart phone GPS modules if the user enabled accurate positioning or approximated POIs with a bounding box (e.g., "The City of Austin" or "Yellowstone National Park") (Yuan et al., 2020).

Numerous studies have used social media data to model human activities. Although these studies may not be directly about activity space modeling, they are valuable for understanding how LBSM can help researchers better study the spatial patterns of human activities. Popular research topics on using LBSM to model spatial activities include, but are not limited to (1) analyzing the clustering and dispersion of human activities; (2) analyzing place semantics and sentiments; and (3) analyzing the interactions on social media and in the geographic space.

First, studies have used location data from social media platforms as a proxy for mobility patterns (Fu et al., 2018; Hawelka et al., 2014; Ilieva & McPhearson, 2018). The results were valuable for understanding urban-level, country-level, and international travel dynamics in the big data era. For instance, Ilieva and McPhearson (2018) reviewed how LBSM data could be used to answer crucial questions in maintaining a sustainable urban system, such as "which parks and green spaces are most popular (Sonter et al., 2016)" and "which metropolitan areas are most likely to suffer from depression (Yang et al., 2015)?"

Second, researchers have analyzed the sentiments and semantics of LBSM user activities and then used the extracted information to enrich the characteristics of geographic places (Doran et al., 2016; Estevez-Ortiz et al., 2016; Mitchell et al., 2013; Sui & Goodchild, 2011). As mentioned in Agnew (2005, p. 84), '... space can be considered as "top-down," defined by powerful actors imposing their control and stories on others. Place can be considered as "bottom-up," representing the outlooks and actions of more typical folks.' Mitchell et al. (2013) conducted a sentiment analysis based on 80 million words generated from a Twitter dataset and mapped the spatial distribution of happiness in the United States. They found the happiest five states, in order, were: Hawaii, Maine, Nevada, Utah, and Vermont, although the differences among all states were not substantial.

Third, researchers in network science have thoroughly studied the structure and evolution of social networks formed on social media sites (Boyd & Ellison, 2007). In contrast, GIScientists focus more on how spatial information plays a factor in the interactions between LBSM users (Jia et al., 2019; Liu et al., 2014). For example, Illenberger et al. (2011) found that the probability of accepting a person as a contact is inversely proportional to the distance between the two users; however, spatial proximity did not seem to impact the topology of social networks. Liu et al. (2014) used a gravity model to explore the distance decay effect in trips extracted from LBSM check-in data.

When explicitly modeling activity spaces, previous research has primarily used LBSM data to understand the differences in activity space indicators among population groups (i.e., similar to the studies based on mobile phone data in Section 3.1) (Hawelka et al., 2014; Li et al., 2013). For instance, Lee et al. (2016) used Twitter data to investigate the differences between activity spaces on weekdays and weekends in Santa Barbara, California. They extracted the convex hulls of Twitter user activity spaces and identified five unique patterns with different shapes and anchor points. Other researchers went one step further and applied LBSM data to studying spatial inequity, social justice, and segregation (Liu et al., 2014; Shelton et al., 2015). A recent work by Wang et al. (2018) found that although residents of primarily Black and Hispanic neighborhoods had similar activity space sizes as residents in advantaged neighborhoods, they were much less likely to visit middle-class neighborhoods, which demonstrated the existence of relative isolation and segregation.

However, compared to georeferenced mobile phone data, LBSM data are not as widely used for activity space modeling due to various reasons. First, the data sampling resolution is much lower than CDRs, let alone MSD (i.e., most people do not post as often on social media as they text or call). Therefore, it requires a longer data collection duration and larger datasets to produce a comprehensive picture of a given user's activity space (Yuan & Wang, 2018). Wang and Yuan (2021) tested the sensitivity of activity space indicators with different sample sizes. Their results showed that although most external and internal indicators approached a stable value with 12 months of data, some indicators could be very unstable and easily distorted, such as the estimated home locations. The spatial sampling rate was often biased too. Figure 4 shows the unbalanced distribution of Twitter check-in data in Austin, TX (Yuan et al., 2020). As can be seen, certain districts, such as the city center, have a higher density of check-ins than other areas.

Second, the sampling rate and demographic bias of LBSM data is concerning to many researchers (Golub & Jackson, 2010; Longley & Adnan, 2016; Longley et al., 2015; Yuan et al., 2020). For example, Pinterest is especially popular among women between the ages of 25 and 34 with average household incomes of $100,000. It is

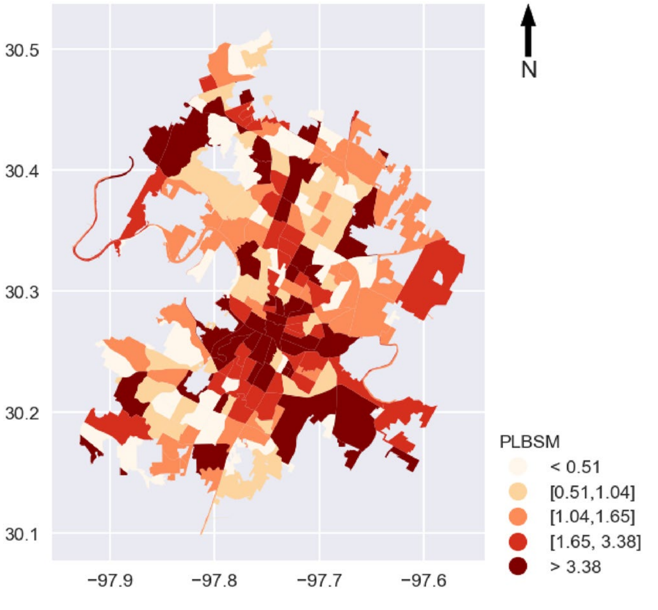

FIGURE 4 Spatial sampling biases of Twitter data in Austin, TX. PLBSM is the number of tweets divided by population and then normalized to the range [0,5]. A darker color (i.e., a higher PLBSM value) indicates that users are more likely to tweet their locations in that district

challenging, sometimes impossible, to accurately assess these demographic biases because most SNS do not validate the profile information provided by users (Yuan et al., 2020).

Third, data authenticity and availability is also a common problem. Fake accounts and bots are inevitable on social networking sites (Dickerson et al., 2014; Gurajala et al., 2015). Because the APIs to access SNS data are provided by the data vendors (i.e., private companies), users have little control over how the samples are generated from a black box or what information is available for research (Gonzalez-Bailon et al., 2014). For example, in June 2019, Twitter announced that it would stop providing the precise geotagging feature (Hu, 2018).

Besides georeferenced mobile phone data and social media data, a few other datasets have been used for activity space modeling. However, they are not as popular as CDRs/MSD or LBSM data for various reasons. One example is studies that recruit participants to wear GPS tracking devices to log their activities (Palmer et al., 2013; Raanan & Shoval, 2014). These studies have similar limitations as traditional travel diaries in that the sample sizes are rather limited, so it is difficult to scale up the analysis. However, the advantage of studies with actively recruited participants is that researchers can obtain more comprehensive and reliable demographic information. Another example is studies that use smart card data to study activity space differences in cities (Gong et al., 2017; Zhang et al., 2021). These datasets are collected through automated transit fare collection systems and can be used to represent the activity patterns of public transit users. For example, Zhang et al. (2021) approximated smart card users' activity spaces as SDEs and found that the spatial extents of elderly and disabled people's activity spaces are smaller than the general population. However, smart card data are limited because they are only available in cities with an advanced public transportation system and obtaining the data requires a close collaboration between researchers and the public sector.

3.4 | Comparing big geo-data in modeling human activity spaces

To sum up, Table 1 provides a simplied summary of the pros and cons of popular datasets for activity space modeing, which can serve as a useful reference for researchers in this field.

Although datasets have their similarities and differences, it is worth noting that most of these big geo-data face one common challenge - how private information can be used ethically and responsibly (Calabrese et al., 2015; Zhang et al., 2016). Countries and regions may have different regulations on how user data, especially telecommunication data, can be collected and used. For example, the European Union published the General Data Protection Regulation

TABLE 1 Comparison of datasets

Data type	Pros	Cons
Travel diaries	• Can include detailed information of participants • Can target interested population groups	• Costly and time-consuming to collect • Often cover limited number of participants and geographic regions • Accuracy relies on participants' input
CDR data	• Easily scalable • Data consistently being collected	• Access to the data requires close partnership with the industry • Low spatial resolution and precision in areas with fewer cell towers • Inconsistent sampling rate in CDR data
LBSM data	• Easily scalable • Easly accessible through APIs	• Potential demographic biases • Low sampling rate • Data authenticity issues (e.g., fake accounts)
Smart card data	• Collected automatically through automated transit fare systems	• Access to the data requires partnership with governmental agencies • Only reflects patterns of public transit use

(GDPR) in 2016, which aims to provide a clear guideline regarding the protection of personal data and privacy issues incurred during data collection, processing, storage, and sharing (European Union, 2016). Data ethics also go beyond what the law regulates. Zook et al. (2017) discussed the best practices of big data research. They highlighted that privacy is not a public/private binary value, so just because something is publicly available and can be legally used, it does not mean that the subsequent use is inherently unproblematic. This is especially important for LBSM studies as users may accidently leave their social media profile public, but it may still be considered inappropriate to look through someone's entire online history (Zook et al., 2017).

4 | CONCLUSIONS

The growing availability of big geo-data provides enormous opportunities for modeling human activity spaces, which has been a classic question in travel behavior analysis for years. In this paper, we first review the commonly used indicators and metrics for human activity space modeling, as well as the strengths and weaknesses of each indicator. Due to the rapid development of ICTs, many new data sources are widely used to study human activities. This paper discusses the progress and challenges of using these new datasets for activity space modeling. In particular, we focus on georeferenced mobile phone data and LBSM data due to their pervasiveness in activity space modeling. We also touch upon studies using other datasets, such as smart card data. The findings can be summarized as follows. First, although big geo-data have various data quality issues (e.g., low spatio-temporal resolutions), they still provide valuation data sources that are easier to obtain and cover a larger geographic area than traditional travel surveys. Second, certain activity space indicators (e.g., estimated home locations) may be more sensitive to sparse datasets than other indicators, so researchers should adjust their data collection strategies based on practical needs. Third, the applications of big geo-data in activity space modeling are not limited to calculating measurements and indicators; instead, researchers have also tried to correlate these measurements with the built environment and social ties to help understand broader research questions, such as social segregation.

A promising future research direction is data fusion techniques that combine big geo-data and traditional data sources. For example, researchers can use travel surveys to validate a subset of the patterns obtained from big geo-data, and then derive a strategy to calibrate the rest of the results. This review paper provides a valuable reference for researchers who are interested in exploring activity space studies using various types of new big geo-data, as well as understanding their strengths and limitations. It also provides a reference for urban planners and policy makers when applying new data sources to plan for a smarter and more efficient urban system in the age of instant access.

ACKNOWLEDGMENTS

Our deep and sincere thanks go to Dr. Caitlin Robinson for her valuable suggestions on the structure of the manuscript. We thank the anonymous reviewers for their constructive comments, which greatly improved the content and clarity of this article. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflict of interest in connection with this article.

ORCID

Yihong Yuan https://orcid.org/0000-0001-6266-9744

REFERENCES

Agnew, J. (2005). Space: Place. In P. J. Cloke & R. J. Johnston (Eds.), *Spaces of geographical thought: Deconstructing human geography's binaries* (pp. 81–96). SAGE Publications.

Ahas, R. (2005). Mobile phones and geography: Social positioning method. In *Power over time-space: Inaugural nordic geographers meeting* (pp. 1–8).

Ahas, R., Aasa, A., Slim, S., Aunap, R., Kalle, H., & Mark, Ü. (2007). Mobile positioning in space-time behaviour studies: Social positioning method experiments in Estonia. *Cartography and Geographic Information Science*, *34*(4), 259–273. https://doi.org/10.1559/152304007782382918

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., & Zook, M. (2015). Everyday space–time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, *29*(11), 1–23. https://doi.org/10.1080/13658816.2015.1063151

Ahas, R., Silm, S., Jarv, O., Saluveer, E., & Tiru, M. (2010). Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, *17*(1), 3–27. https://doi.org/10.1080/10630731003597306

Bachi, R. (1963). Standard distance measures and related methods for spatial analysis. *Papers - Regional Science Association*, *10*(1), 83–132. https://doi.org/10.1111/j.1435-5597.1962.tb00872.x

Bayir, M. A., Demirbas, M., & Eagle, N. (2009). Discovering spatiotemporal mobility profiles of cellphone users. In *2009 IEEE international symposium on a world of wireless* (pp. 1–9). Mobile and Multimedia Networks & Workshops.

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *Epj Data Science*, *4*(1), 10. https://doi.org/10.1140/epjds/s13688-015-0046-0

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*(6264), 1073–1076. https://doi.org/10.1126/science.aac4420

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brown, L. A., & Moore, E. G. (1970). The intra-urban migration process: A perspective. *Geografiska Annaler*, *52*, 1–13. https://doi.org/10.1080/04353684.1970.11879340

Buliung, R. N., & Kanaroglou, P. S. (2006). A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography*, *14*(1), 35–51. https://doi.org/10.1016/j.jtrangeo.2004.10.008

Calabrese, F., Ferrari, L., & Blondel, V. D. (2015). Urban sensing using mobile phone network data: A survey of research. *ACM Computing Surveys*, *47*(2), 1–20. https://doi.org/10.1145/2655691

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS One*, *6*(7), e20814. https://doi.org/10.1371/journal.pone.0020814

Chen, C., Bian, L., & Ma, J. T. (2014). From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies*, *46*, 326–337. https://doi.org/10.1016/j.trc.2014.07.001

Dickerson, J. P., Kagan, V., & Subrahmanian, V. S. (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *2014 proceedings of the ieee/acm international conference on advances in social networks analysis and mining (asonam 2014)* (pp. 620–627).

Dijst, M. (1999). Two-earner families and their action spaces: A case study of two Dutch communities. *Geojournal*, *48*(3), 195–206. https://doi.org/10.1023/a:1007031809319

Doran, D., Severin, K., Gokhale, S., & Dagnino, A. (2016). Social media enabled human sensing for smart cities. *AI Communications*, *29*(1), 57–75. https://doi.org/10.3233/aic-150683

Estevez-Ortiz, F. J., Garcia-Jimenez, A., & Glosekotter, P. (2016). An application of people's sentiment from social media to smart cities. *Profesional De La Informacion*, *25*(6), 851–858. https://doi.org/10.3145/epi.2016.nov.02

European Union. (2016). Regulation (eu) 2016/679 of the european parliament and of the council.

Fohringer, J., Dransch, D., Kreibich, H., & Schroter, K. (2015). Social media as an information source for rapid flood inundation mapping. *Natural Hazards and Earth System Sciences*, *15*(12), 2725–2738. https://doi.org/10.5194/nhess-15-2725-2015

Frias-Martinez, V., Soguero-Ruiz, C., Frias-Martinez, E., & Josephidou, M. (2013). Forecasting socioeconomic trends with cell phone records. In *Proceedings of the 3rd ACM symposium on computing for development*. Association for Computing Machinery. Article 15.

Fu, C., McKenzie, G., Frias-Martinez, V., & Stewart, K. (2018). Identifying spatiotemporal urban activities through linguistic signatures. *Computers, Environment and Urban Systems*, *72*, 25–37. https://doi.org/10.1016/j.compenvurbsys.2018.07.003

Gesler, W. M., & Meade, M. S. (1988). Locational and population factors in health care-seeking behavior in Savannah, Georgia. *Health Services Research*, *23*.

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographic perspective*. Guilford Press.

Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149. https://doi.org/10.1257/mic.2.1.112

Gong, Y. X., Lin, Y. Y., & Duan, Z. Y. (2017). Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Computers, Environment and Urban Systems*, *64*, 169–183. https://doi.org/10.1016/j.compenvurbsys.2017.02.003

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. https://doi.org/10.1038/nature06958

Gonzalez-Bailon, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, *38*, 16–27. https://doi.org/10.1016/j.socnet.2014.01.004

Gurajala, S., White, J. S., Hudson, B., & Matthews, J. N. (2015). Fake twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach. In *Proceedings of the 2015 international conference on social media & society* (pp. 1–7). ACM.

Hägerstrand, T. (1970). What about people in regional science? *Papers - Regional Science Association*, *24*(1), 7–21. https://doi.org/10.1007/bf01936872

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*(3), 260–271. https://doi.org/10.1080/15230406.2014.890072

Horton, F., & Reynolds, D. R. (1971). Effects of urban spatial structure on individual behaviour. *Economic Geography*, *47*(1), 36–48. https://doi.org/10.2307/143224

Hu, L., Li, Z., & Ye, X. (2020). Delineating and modeling activity space using geotagged social media data. *Cartography and Geographic Information Science*, *47*(3), 277–288. https://doi.org/10.1080/15230406.2019.1705187

Hu, Y. (2018). Geo-text data and data-driven geospatial semantics. *Geography Compass*, *12*(11), e12404. https://doi.org/10.1111/gec3.12404

Huang, X., Lu, J. Y., Gao, S., Wang, S. C., Liu, Z. W., & Wei, H. X. (2022). Staying at home is a privilege: Evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. *Annals of the American Association of Geographers*, *112*(1), 286–305. https://doi.org/10.1080/24694452.2021.1904819

Ilieva, R. T., & McPhearson, T. (2018). Social-media data for urban sustainability. *Nature Sustainability*, *1*(10), 553–565. https://doi.org/10.1038/s41893-018-0153-6

Illenberger, J., Kowald, M., Axhausen, K. W., & Nagel, K. (2011). Insights into a spatially embedded social network from a large-scale snowball sample. *The European Physical Journal B*, *84*(4), 549–561. https://doi.org/10.1140/epjb/e2011-10872-0

International Telecommunication Union. (2021). Facts and figures.

Isaacman, S., Becker, R., Caceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. *Pervasive Computing*, *6696*, 133–151.

Janecek, A., Valerio, D., Hummel, K. A., Ricciato, F., & Hlavacs, H. (2015). The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring. *IEEE Transactions on Intelligent Transportation Systems*, *16*(5), 2551–2572. https://doi.org/10.1109/tits.2015.2413215

Jarv, O., Muurisepp, K., Ahas, R., Derudder, B., & Witlox, F. (2015). Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. *Urban Studies*, *52*(14), 2680–2698. https://doi.org/10.1177/0042098014550459

Jia, T., Yu, X. S., Shi, W. Z., Liu, X. T., Li, X., & Xu, Y. (2019). Detecting the regional delineation from a network of social media user interactions with spatial constraint: A case study of Shenzhen, China. *Physica A: Statistical Mechanics and its Applications*, 531. https://doi.org/10.1016/j.physa.2019.121719

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *Ieee Transactions on Big Data*, *3*(2), 208–219. https://doi.org/10.1109/tbdata.2016.2631141

Jones, P., Koppelman, F., & Orfueil, J. P. (1990). Activity analysis: State-of-the-art and future. In P. Jones (Ed.), *Developments in dynamic and activity-based approaches to travel analysis* (pp. 34–55). Avebury.

Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., & Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *2010 18th international conference on geoinformatics* (pp. 1–7).

Kwan, M. P. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*, *30*(3), 191–216. https://doi.org/10.1111/j.1538-4632.1998.tb00396.x

Kwan, M. P. (1999). Gender and individual access to urban opportunities: A study using space–time measures. *The Professional Geographer*, *51*(2), 211–227. https://doi.org/10.1111/0033-0124.00158

Lee, J. H., Davis, A. W., Yoon, S. Y., & Goulias, K. G. (2016). Activity space estimation with longitudinal observations of social media data. *Transportation*, *43*(6), 955–977. https://doi.org/10.1007/s11116-016-9719-1

Leo, Y., Fleury, E., Alvarez-Hamelin, J. I., Sarraute, C., & Karsai, M. (2016). Socioeconomic correlations and stratification in social-communication networks. *Journal of The Royal Society Interface*, *13*(125), 20160598. https://doi.org/10.1098/rsif.2016.0598

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, *40*(2), 61–77. https://doi.org/10.1080/15230406.2013.777139

Liu, Y., Sui, Z. W., Kang, C. G., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social Media check-in data. *PLoS One*, *9*(1), e86026. https://doi.org/10.1371/journal.pone.0086026

Longley, P. A., & Adnan, M. (2016). Geo-temporal twitter demographics. *International Journal of Geographical Information Science*, *30*(2), 369–389. https://doi.org/10.1080/13658816.2015.1089441

Longley, P. A., Adnan, M., & Lansley, G. (2015). The geotemporal demographics of Twitter usage. *Environment and Planning A*, *47*(2), 465–484. https://doi.org/10.1068/a130122p

Mancosu, M., & Bobba, G. (2019). Using deep-learning algorithms to derive basic characteristics of social media users: The Brexit campaign as a case study. *PLoS One*, *14*(1), e0211013. https://doi.org/10.1371/journal.pone.0211013

Mazey, M. E. (1981). The effect of a physio-political barrier upon urban activity space. *Ohio Journal of Science*, *81*, 212–217.

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One*, *8*(5), e64417. https://doi.org/10.1371/journal.pone.0064417

Newsome, T. H., Walcott, W. A., & Smith, P. D. (1998). Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation*, *25*(4), 357–377. https://doi.org/10.1023/a:1005082827030

Nyhan, M., Grauwin, S., Britter, R., Misstear, B., McNabola, A., Laden, F., Barrett, S. R. H., & Ratti, C. (2016). "Exposure track" the impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. *Environmental Science & Technology*, *50*(17), 9671–9681. https://doi.org/10.1021/acs.est.6b02385

Palmer, J. R. B., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E., & Li, K. (2013). New approaches to human mobility: Using mobile phones for demographic research. *Demography*, *50*(3), 1105–1128. https://doi.org/10.1007/s13524-012-0175-z

Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., & Giannotti, F. (2016). An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, *2*(1–2), 75–92. https://doi.org/10.1007/s41060-016-0013-2

Pendyala, R. M., Goulias, K. G., & Kitamura, R. (1991). Impact of telecommuting on spatial and temporal patterns of household travel. *Transportation*, *18*(4), 383–409. https://doi.org/10.1007/bf00186566

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In A. A. Salah, T. Gevers, N. Sebe, & A. Vinciarelli (Eds.), *Human behavior understanding* (pp. 14–25).

Phithakkitnukoon, S., Smoreda, Z., & Olivier, P. (2012). Socio-geography of human mobility: A study using longitudinal mobile phone data. *PLoS One*, *7*(6), e39253. https://doi.org/10.1371/journal.pone.0039253

Raanan, M. G., & Shoval, N. (2014). Mental maps compared to actual spatial behavior using GPS data: A new method for investigating segregation in cities. *Cities*, *36*, 28–40. https://doi.org/10.1016/j.cities.2013.09.003

Raskind, I. G., Kegler, M. C., Girard, A. W., Dunlop, A. L., & Kramer, M. R. (2020). An activity space approach to understanding how food access is associated with dietary intake and BMI among urban, low-income African American women. *Health & Place*, *66*, 102458. https://doi.org/10.1016/j.healthplace.2020.102458

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, *33*(5), 727–748. https://doi.org/10.1068/b32047

Ren, M., Park, S., Xu, Y., Huang, X., Zou, L., Wong, M. S., & Koh, S.-Y. (2022). Impact of the COVID-19 pandemic on travel behavior: A case study of domestic inbound travelers in Jeju, Korea. *Tourism Management*, *92*, 104533. https://doi.org/10.1016/j.tourman.2022.104533

Roick, O., & Heuser, S. (2013). Location based social networks - definition, current state of the art and research agenda. *Transactions in GIS*, *17*, 763–784. https://doi.org/10.1111/tgis.12032

Schönfelder, S., & Axhausen, K. W. (2002). Measuring the size and structure of human activity spaces the longitudinal perspective. *Arbeitsbericht Verkehrs- und Raumplanung*, 135.

Schönfelder, S., & Axhausen, K. W. (2003). Activity spaces: Measures of social exclusion? *Transport Policy*, *10*(4), 273–286. https://doi.org/10.1016/j.tranpol.2003.07.002

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, *142*, 198–211. https://doi.org/10.1016/j.landurbplan.2015.02.020

Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M., & Arcury, T. A. (2005a). A suite of methods for representing activity space in a healthcare accessibility study. *International Journal of Health Geographics*, *4*(1), 24. https://doi.org/10.1186/1476-072x-4-24

Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M., & Arcury, T. A. (2005b). A suite of methods for representing activity space in a healthcare accessibility study. *International Journal of Health Geographics*, *4*, 1–21. https://doi.org/10.1186/1476-072x-4-24

Shi, L., Wu, L., Chi, G. H., & Liu, Y. (2016). Geographical impacts on social networks from perspectives of space and place: An empirical study using mobile phone data. *Journal of Geographical Systems*, *18*(4), 359–376. https://doi.org/10.1007/s10109-016-0236-8

Silm, S., & Ahas, R. (2014). Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers*, *104*(3), 542–559. https://doi.org/10.1080/00045608.2014.892362

Song, C. M., Qu, Z. H., Blumm, N., & Barabasi, A. L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170

Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS One*, *11*(9), e0162372. https://doi.org/10.1371/journal.pone.0162372

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748. https://doi.org/10.1080/13658816.2011.604636

Thériault, M., Claramunt, C., & Villeneuve, P. Y. (1999). A spatio-temporal taxonomy for the representation of spatial set behaviours. In M. H. Böhlen, C. S. Jensen, & M. O. Scholl (Eds.), *Spatio-temporal database management* (pp. 1–18). Springer Berlin Heidelberg.

Toole, J. L., Herrera-Yaque, C., Schneider, C. M., & Gonzalez, M. C. (2015). Coupling human mobility and social ties. *Journal of The Royal Society Interface*, *12*(105), 20141128. https://doi.org/10.1098/rsif.2014.1128

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences*, *115*(30), 7735–7740. https://doi.org/10.1073/pnas.1802537115

Wang, X., & Yuan, Y. (2021). Modeling user activity space from location-based social media: A case study of Weibo. *The Professional Geographer*, *73*(1), 96–114. https://doi.org/10.1080/00330124.2020.1803090

Wang, Y., Kang, C., Bettencourt, L. M., Liu, Y., & Andris, C. (2015). Linked activity spaces: Embedding social networks in urban space. In *Computational approaches for urban environments* (pp. 313–336). Springer.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of The Royal Society Interface*, *10*(81), 20120986. https://doi.org/10.1098/rsif.2012.0986

Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS One*, *9*(5), e97010. https://doi.org/10.1371/journal.pone.0097010

Xu, W. (2021). The contingency of neighbourhood diversity: Variation of social context using mobile phone application data. *Urban Studies*, *59*(4), 851–869. https://doi.org/10.1177/00420980211019637

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2017). How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS*, *21*(3), 468–487. https://doi.org/10.1111/tgis.12285

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems*, *72*, 51–67. https://doi.org/10.1016/j.compenvurbsys.2018.04.001

Xu, Y., Belyi, A., Santi, P., & Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *Journal of The Royal Society Interface*, *16*(160), 20190536. https://doi.org/10.1098/rsif.2019.0536

Xu, Y., Li, J. Y., Xue, J. Y., Park, S., & Li, Q. Q. (2021). Tourism geography through the lens of time use: A computational framework using fine-grained mobile phone data. *Annals of the American Association of Geographers*, *111*(5), 1420–1444. https://doi.org/10.1080/24694452.2020.1812372

Xu, Y., Li, X. Y., Shaw, S. L., Lu, F., Yin, L., & Chen, B. Y. (2020). Effects of data preprocessing methods on addressing location uncertainty in mobile signaling data. *Annals of the American Association of Geographers*, *111*(2), 515–539. https://doi.org/10.1080/24694452.2020.1773232

Xu, Y., Shaw, S. L., Fang, Z. X., & Yin, L. (2016). Estimating potential demand of bicycle trips from mobile phone data-an anchor-point based approach. *ISPRS International Journal of Geo-Information*, *5*(8), 131. https://doi.org/10.3390/ijgi5080131

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation*, *42*(4), 625–646. https://doi.org/10.1007/s11116-015-9597-y

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z., & Li, Q. (2016). Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, *106*, 489–502.

Xu, Y., Xue, J. Y., Park, S., & Yue, Y. (2021). Towards a multidimensional view of tourist mobility patterns in cities: A mobile phone data perspective. *Computers, Environment and Urban Systems*, *86*, 101593. https://doi.org/10.1016/j.compenvurbsys.2020.101593

Yamamoto, T., & Kitamura, R. (1999). An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non-working days. *Transportation*, *26*, 211–230.

Yan, L. X., Wang, D., Zhang, S. W., & Xie, D. C. (2019). Evaluating the multi-scale patterns of jobs-residence balance and commuting time-cost using cellular signaling data: A case study in Shanghai. *Transportation*, *46*(3), 777–792. https://doi.org/10.1007/s11116-018-9894-3

Yang, W., Mu, L., & Shen, Y. (2015). Effect of climate and seasonality on depressed mood among twitter users. *Applied Geography*, *63*, 184–191. https://doi.org/10.1016/j.apgeog.2015.06.017

Yuan, Y., Lu, Y., Chow, T. E., Ye, C., Alyaqout, A., & Liu, Y. (2020). The missing parts from social media–enabled smart cities: Who, where, when, and what? *Annals of the American Association of Geographers*, *110*(2), 462–475. https://doi.org/10.1080/24694452.2019.1631144

Yuan, Y., & Raubal, M. (2016). Analyzing the distribution of human activity space from mobile phone usage: An individual and urban-oriented study. *International Journal of Geographical Information Science*, *30*(8), 1594–1621. https://doi.org/10.1080/13658816.2016.1143555

Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior - a case study of Harbin, China. *Computers, Environment and Urban Systems*, *36*(2), 118–130. https://doi.org/10.1016/j.compenvurbsys.2011.07.003

Yuan, Y., & Wang, X. (2018). Exploring the effectiveness of location-based social media in modeling user activity space: A case study of Weibo. *Transactions in GIS*, *22*(4), 930–957. https://doi.org/10.1111/tgis.12450

Yuan, Y., Wei, G., & Lu, Y. (2018). Evaluating gender representativeness of location-based social media: A case study of Weibo. *Annals of GIS*, *24*(3), 163–176. https://doi.org/10.1080/19475683.2018.1471518

Yuill, R. S. (1971). The standard deviational ellipse: An updated tool for spatial description. *Geografiska Annaler*, *53B*.

Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C., & Stokes, C. (2011). Activity space environment and dietary and physical activity behaviors: A pilot study. *Health & Place*, *17*(5), 1150–1161. https://doi.org/10.1016/j.healthplace.2011.05.001

Zhang, S. Q., Yang, Y., Zhen, F., Lobsang, T., & Li, Z. X. (2021). Understanding the travel behaviors and activity patterns of the vulnerable population using smart card data: An activity space-based approach. *Journal of Transport Geography*, *90*, 102938. https://doi.org/10.1016/j.jtrangeo.2020.102938

Zhang, Y., Chen, Q. J., & Zhong, S. (2016). Privacy-preserving data aggregation in mobile phone sensing. *IEEE Transactions on Information Forensics and Security*, *11*(5), 980–992. https://doi.org/10.1109/tifs.2016.2515513

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology*, *13*(3), e1005399. https://doi.org/10.1371/journal.pcbi.1005399

AUTHOR BIOGRAPHIES

Dr. Yihong Yuan received her Ph.D. in Geography and M.A in Statistics from the University of California, Santa Barbara, and her B.S. in Geographic Information Systems from Peking University, China. She is currently an Associate Professor of Geography and Environmental Studies at Texas State University. Dr. Yuan's research focuses on big geo-data analytics and spatial-temporal knowledge discovery. She is particularly interested in extracting human activity patterns from multiple data sources, such as telecommunication devices and social networking sites.

Dr. Yang Xu received his Ph.D. in Geography from the University of Tennessee in Knoxville. He is currently an Assistant Professor in the Department of Land Surveying and Geo-Informatics at the Hong Kong Polytechnic University, where he directs the program of "Master of Science in Urban Informatics and Smart Cities." His research interests include GIScience, human mobility, and urban informatics.

How to cite this article: Yuan, Y., & Xu, Y. (2022). Modeling activity spaces using big geo-data: Progress and challenges. *Geography Compass*, e12663. https://doi.org/10.1111/gec3.12663