Urban Forestry & Urban Greening 31 (2018) 130–141

Contents lists available at ScienceDirect

Urban Forestry & Urban Greening

journal homepage: www.elsevier.com/locate/ufug

Emerging social media data on measuring urban park use

Yiyong Chen[a], Xiaoping Liu[b,*], Wenxiu Gao[a], Raymond Yu Wang[c], Yun Li[a], Wei Tu[a]

[a] Shenzhen Key Laboratory of Built Environment Optimization, School of Architecture & Urban Planning, Shenzhen University, Shenzhen 518060, China
[b] School of Geography and Planning, Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou 510275, China
[c] Center for Chinese Public Administration Research, School of Government, Sun Yat-sen University, Guangzhou 510275, China

ARTICLE INFO

Keywords:
Green space
RTUD data
Shenzhen
Time-spatial distribution
User density

ABSTRACT

Green urban infrastructure, which serves the interests of both human and nature, are considered as essential assets of urban residents. However, measuring the use of green space has been problematic. Because, most previously used data on measuring green space use are self-reported or collected via social surveys, which not only include limited samples, but also are always subjective, costly, and laborious. This study integrates sensor and positioning technologies and measures the use of green space from an emerging big data perspective. The hourly real-time Tencent user density (RTUD) data from social media are used to analyze the time-spatial distribution of urban park users. RTUD data, park attributes, and surrounding landscape features are incorporated into ArcGIS for spatial analysis. A group of linear regression models is constructed to determine factors that may be associated with the user density of urban parks. The total accumulated number of observed users is 3.25 million in 686 urban parks of Shenzhen in two typical sunny days – a work day and a rest day. Without costly and laborious field investigation, based on RTUD data, we conduct a city wide analysis regarding all parks. The proposed method is proved to be suitable for measuring urban green space use at city-level or even large scale. Results show that park user density is relatively high in well-developed areas and community parks. Park attributes and surrounding landscape features are significantly associated to park use. The findings of this study can help policy makers optimize the construction and maintenance of urban parks.

1. Introduction

Green urban infrastructure, such as parks, gardens, greenways, forests, wetlands, agricultural fields, green roofs, green walls, and other kinds of urban green space, are considered essential assets of urban residents (Benedict and Mcmahon, 2006; Goličnik Marušić, 2015). The provision of green space has received increasing scholarly attention in the field of urban and landscape planning. Numerous studies highlighted the importance of enhancing the use of green space given their positive effect on the health and well-being of urban residents. Green space, a key element of built environment in urban areas, could encourage moderate-to-vigorous physical activities among residents (Bowler et al., 2010; Coutts et al., 2013; Kaczynski and Henderson, 2008), and provide a number of health benefits, such as reduce stress (Jansson et al., 2013; Jiang et al., 2014; Roe et al., 2013) and increase positive emotions (Ulrich et al., 1991), restore attention (Kaplan, 1993), reduce obesity and cardiovascular disease (Sallis et al., 2012), improved self-reported health (Maas et al., 2006).

Dynamic factors and their complex interactions affect the use of urban green spaces such as size (Giles-Corti et al., 2005; Tamosiunas et al., 2014), the presence of facilities (Kaczynski and Potwarka, 2009; Sugiyama et al., 2015), accessibility (Koohsari et al., 2013; Schipperijn et al., 2010a; Tzoulas et al., 2007), and the surrounding landscape features (Chen et al., 2016; Coutts, 2009). In most studies, distance is considered to be the most important factor related to green space use (Koohsari et al., 2015; Zhang et al., 2011). Specifically, close proximity to green spaces is associated with increased use (Akpinar, 2016; Schipperijn et al., 2010b). However, in some cases, distance to green space is found not a limiting factor when considering neighbor landscape features (Schipperijn et al., 2010a). The use of parks is also influenced by aesthetic features, presence of amenities (Goličnik and Ward Thompson, 2010; Holman et al., 1996). Based on these factors, many countries formulated various policies and guidelines that guide the planning and design of public green spaces with the aim of promoting the use of green space. Landscape planning and the design of green space, which usually cover large areas with complex components and functions, have become an important task in the field of landscape and urban planning (Koohsari et al., 2015).

* Corresponding author.
E-mail address: liuxp3@mail.sysu.edu.cn (X. Liu).

https://doi.org/10.1016/j.ufug.2018.02.005
Received 8 November 2017; Received in revised form 2 February 2018; Accepted 12 February 2018
Available online 02 March 2018
1618-8667/

Table 1
Pros and cons of different sources of data regarding green space use.

Item	In-situ observed user data	Self-reported data	GPS and VGI big data	RTUD data
Data source	In-situ Observe, field investigation	Questionnaire, interview	Electronic GPS devices, social media	Electronic GPS devices with Tencent Apps
Reliability	Depend on sampling method and socio-demographic attributes of respondents	Depend on sampling method, sampling scale and case chosen	Data from volunteer's personnel sharing; depend on the spatial & temporal resolution and scale of data; historical data available from social media	Data from automatically statistics of Tencent Co., publicly available from Tencent website; only real time data, no historical data; now only apply to Mainland China
Validity	Rely on sampling scale, investigation design and modeling method; but are valuable on measuring relation between activity and public space	Rely on sampling scale, questionnaire design and modeling method, but are valuable on studies of perception, attitude, requirements	Rely on source devices and modeling method, VGI data are only representative for volunteers; can record and measure user's position and movement	Hourly user data at city scale with high spatial resolution, which counts the users in green space using Tencent Apps (In 2016, above 70% of the total population in China use Tencent Apps); valid only on statistics of user count; not applicable for small open spaces (25 m spatial resolution)
Scale of the sample	Limited number of samples, usually dozens of green spaces, hundreds of users	Limited number of samples, usually hundreds or thousands of users	Millions of users related to thousands of green spaces; VGI scale differs, from thousands to millions	Billions of users in China and millions in our case in each hour; wide coverage, high spatial and temporal resolution
Cost of data collect	Need a lot of social investigation; costly and laborious	Need a lot of labor to collect data; costly and laborious	Data from electronic devices and open sources; cheap and effortless	Crawl directly from website; cheap and effortless; high technical requirements
Level of detail of the data	Include in-situ behavior or activity of user and its relation with site	Detail socio-demographic attributes and attitude, perception of users	Include the attributes such as overall spatial distribution and dynamics, without personnel attributes	Show spatial distribution of users; but without personnel attributes, perception, activities, interaction or attitudes

However, measuring the use of green space has been problematic. Most of above mentioned studies depend on direct site observations or social investigation of parks. The field investigation requires tremendous time and funding. Moreover, the measurement of the frequency and duration of green space use mainly depends on self-reported or observed data via social surveys (Hillsdon et al., 2006; Schipperijn et al., 2010a), which have several shortcomings (see Table 1): (1) the amount of collected samples is limited, and data reliability highly rely on survey methods. The limited sample usually leads to a deviation between the total population and the surveyed samples (Cohen et al., 2016). In addition, some studies adopted the observation data of physical activities in green spaces (Chen et al., 2016; v Janowsky and Becker, 2003), but only limited green space can be observed given the time constraint. For example, a recently national study of Neighborhood Parks in the United States adopted an average sample of 6 neighborhood parks in each of the 25 major cities (Cohen et al., 2016). Thus, limited data can be obtained. (2) The validity and reliability of the self-reported data is usually questionable; because the accuracy of self-reported data highly depends on socio-demographic attributes of the respondents, whose objective behavior may deviate from their subjective description (Jiang et al., 2016). (3) Traditional data collecting method requires considerable amount of work on data collection (Rigolon and Nemeth, 2016). In some studies, infrared counters are used to measure the number of visitors who automatically enter green spaces (Lindsey and Nguyen, 2004), but this method is also costly and laborious, and can only be applied to measure the crossing flow of visitors of green spaces with certain entrances, such as gated parks and linear greenways. Nonetheless, research methods and data collected differs in many ways from surveys, which highly rely on validity and reliability of observation and investigation tools (Fink, 2002; Veal, 2011).

Given the limitations of traditional social surveys, many studies attempted to measure urban green space use with new methods and data. For instance, geospatial data such as geo-referenced photos from Panoramio and Flickr (Alivand and Hochmair, 2017; Levin et al., 2017) or other social medias (García-Palomares et al., 2015) are used to describe park attributes and use. These data allow researchers to study human social and physical behavior regarding park use from an emerging big-data perspective. Nonetheless, none of these data sources could measure park use at city-level on daily or hourly intervals. Some questions are still unclear: how many visitors can be observed in all kinds of urban parks in a day? What are the hourly and daily differences of park use? What are the differences among parks of different district, location and category? What are the factors influencing the user density of parks from a city-level prospective? Answers for these questions could offer good suggestions for policy-makers and landscape planners to increase urban park use.

The emergence of location based social media big data, such as real-time Tencent user density (RTUD) data, provides an alternative on measuring the time-spatial distribution of use of green space. The RTUD, which is provided by Tencent (http://www.qq.com), contains real-time hourly numbers of mobile device users who use Tencent applications, such as Tencent Mobile App QQ, WeChat, Soso Maps, and other mobile applications that provide Location-Based Service (LBS) using GPS devices with a spatial accuracy level of about 10 m (Yao et al., 2017). The Tencent Company records users' location when they use LBS Apps, and counts the amount of users in each pixel unit, which resulted in the RTUD data. The RTUD data is shown at the publicly available Tencent website at a spatial resolution of 25 m, while every 25 m*25 m pixel contains a value representing count of users inside this area. Previous studies show that the RTUD data could accurately reflect activity distribution and can be well used in urban and transportation studies (Niu et al., 2017; Yao et al., 2017; Zhang et al., 2014). In one explorative study RTUD data are integrated with taxi data and high-resolution Quickbird images to infer building functions, which shows an overall detection rate of 72.22% and an accuracy rate of above 65%