INVESTIGATING THE ROLE OF SOCIAL MEDIA AND SMART DEVICE APPLICATIONS IN UNDERSTANDING HUMAN-ENVIRONMENT RELATIONSHIPS IN URBAN GREEN SPACES

by

HELEN VICTORIA ROBERTS

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Geography, Earth and Environmental Sciences
College of Life and Environmental Sciences
University of Birmingham

February 2018

Abstract

Urban green spaces are integral components of urban landscapes and the cultural ecosystem services afforded to human populations by these green spaces are of particular relevance to human and societal well-being. Urban green spaces provide opportunities for human interaction, physical activity and recreation, stress alleviation and mental restoration, economic opportunity, cultural activities and interactions with nature. To understand how these benefits are received by human populations it is vital to understand when and how individuals interact with urban green spaces. The rapid development and uptake of technologies such as smart phones, social networks and apps provides new opportunity to investigate the human interactions occurring in urban green spaces. Using the city of Birmingham as a case study, this thesis aims (i) to *demonstrate* the utility of data obtained from smart device enabled platforms (social networks and apps) in understanding socio-ecological interactions in urban areas and (ii) to *evaluate* the utility of these data sources for researchers and policy makers. The successful identification of a range of socio-ecological interaction suggest these data sources provide a viable method of investigating such interactions; however, there remain a number of limitations to consider to ensure they are employed appropriately in research contexts.

Acknowledgements

The development and production of this thesis was made considerably easier thanks to a number of individuals. I am indebted to my supervisors Professor Jon Sadler and Professor Lee Chapman for their guidance, enthusiasm and optimism, without which the research and production of this thesis would have seemed an impossible task. I would also like to extend thanks to Dr Phil Jones for providing support and encouragement to engage with opportunities beyond those which are standard to PhD research.

I gratefully acknowledge the contributions of Dr Bernd Resch and his colleagues who provided the semi-supervised learning algorithm presented in Chapter Six. His expertise and commitment to the project were significant influences in the success of the resulting paper from this chapter. I would also like to thank Simon Bell for obtaining and providing the Netatmo datasets used in Chapter Four.

I am grateful to my colleagues in the 225 postgraduate office for their support, advice and friendship. Thank you for providing a welcoming and enjoyable working environment. Finally, I would like to thank my family for their constant support and encouragement in my endeavours, and for believing in me when I struggle to believe in myself.

This research was made possible by financial support from an Engineering and Physical Science Research Council (EPSRC) studentship (EP/L504907/1).

Table of Contents

List of figures:

Table of tables:

Chapter 1. Introduction

1.1 The expanding urban landscape

The world is becoming increasingly urbanised and increasing amounts of the land surface are becoming part of urban agglomerations (Seto et al., 2011). The majority of the global population now resides in cities (Wu, 2008; UNHABITAT, 2016) which has implications for these urban human collectives in terms of health and well-being. The density of the world's urban areas will continue to increase during this century (Burton, 2002; Coutts et al., 2006; Irwin and Bockstael, 2007) therefore an appreciation of how urban landscapes impact on human populations has become increasingly important (Grimm et al., 2008).

As urban areas increase in extent, their physical form becomes increasingly heterogeneous (Wu et al., 2014) and there is a continual need to study them at a high spatial resolution (Cadenasso et al., 2007). Measuring such variation is essential in supporting urban research and decision making within the social sphere, and methods need continual advancement in order to best capture the nature of the urban landscape (Hale, 2015). Urban green spaces are increasingly cited as influential for the health and well-being of city populations (Chiesura, 2004). Their location, size, predominant vegetation type and density may be important in understanding how the benefits they provide are spatialized across the cityscape (Diaz and Cabido, 2001; Kremen, 2005). Aside from the physical characteristics of urban green spaces, their use by humans is of key significance in understanding how the benefits of ecosystems are transferred to and received by the population (Daniel et al., 2012).

1.2 Ecosystem services and specifically cultural ecosystem services

The broadly defined term of ecosystem services encapsulates all the potential benefits that human populations receive from the ecosystems that surround them and is an essential conceptualisation that has helped advance socio-ecological systems research (Pickett et al., 1997; Tzoulas et al., 2007). As a collective concept, ecosystem services define the provisioning, supporting, regulating and cultural benefits offered to human populations by ecosystem (Costanza et al., 1997; Daily, 1997; Ehrlich and Ehrlich, 1981; MEA, 2005).

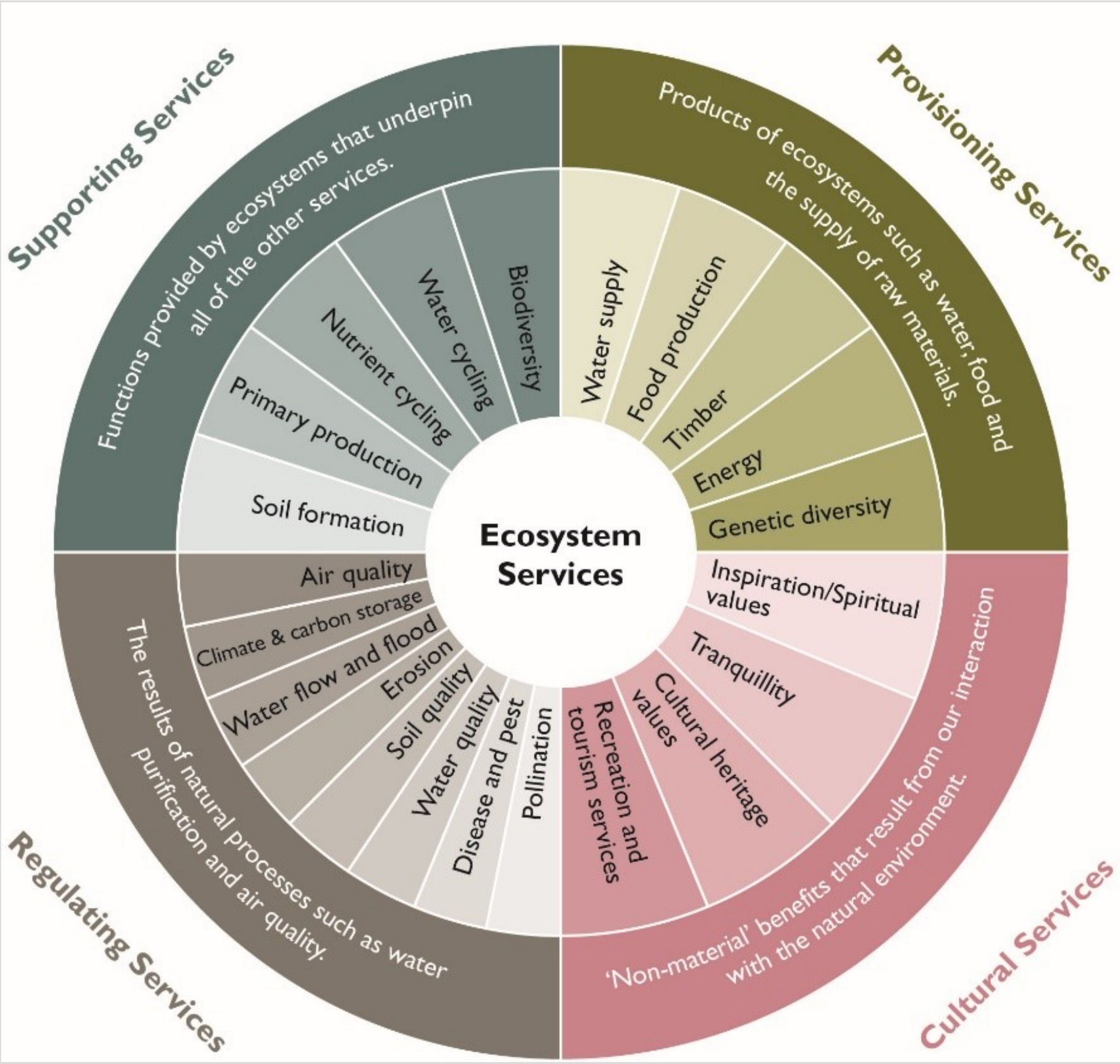


Figure 1.1 *Ecosystem services provided to human populations by natural ecosystems (SDNPA, 2013).*

This thesis is primarily concerned with 'cultural ecosystem services' and specifically focuses on the provision of opportunity for engagement with recreational and physical activities and the psychological benefits that natural ecosystems provide (MEA, 2005).

There is increasing evidence that urban green spaces provide restorative and preventative health benefits (Maas et al., 2006, van den Berg et al., 2010), a notion central to the framework of this thesis. Given the increasing levels of obesity, cardiovascular disease and mental illness reported in England (NHS, 2016; ONS, 2016), the role of urban green spaces in providing low cost mitigation to these problems through their ability to maintain and improve the physical and mental health of human urban populations has become increasingly significant for both researchers and policy makers (DEFRA, 2017). Despite this, cultural ecosystem services have remained little studied in urban areas given the challenges faced in by researchers in attempts to quantify their delivery and receipt to and by human populations (Daniel et al., 2012).

The intangibility and lack of economic quantification of cultural ecosystem services has hampered their assessment and, as a result, there remains a lack of integration between the ecosystem service framework and planning system (de Groot et al., 2005; Kabisch et al., 2015): and despite them offering a range of benefits which are recognised by decision makers, their inclusion in policy is often sacrificed for economic or ecological reasons given that these ecosystem services are somewhat easier to quantify (Milcu et al., 2013).

It is encouraging, however, that the planning literature acknowledges the benefits of ecosystems to human and societal well-being (Egoh et al., 2007; Bennett et al., 2009), indicating that planners and decision makers are familiar with the concept of ecosystem services and are willing to engage with it (Kabisch et al., 2015). Research must therefore provide the necessary tools for them to do so, and developing new means of assessing cultural ecosystem value is essential in this process given that while engaging a wide variety of disciplines, theoretical and methodological approaches, cultural ecosystem research is yet to reach a satisfactory understanding of the benefits provided by ecosystems to human populations (de Groot et al., 2005). The multitude of perspectives through which cultural ecosystem services can be viewed reflects the development of the relatively new area of research that lacks a well-established research framework.

Establishing ways to value and quantify the benefits that urban populations receive from the natural ecosystems that surround them will provide information that can be used to inform when and how these benefits are received, promoting an evidence-based approach for creating future spaces which maximise the potential for ecosystem service receipt by urban citizens. Indeed, current government rhetoric identifies “a lack of evidence specifically designed to inform the development of policy and interventions” in attempts to integrate management of human well-being with natural environments (DEFRA, 2017, p. 30).

Incorporating cultural services into ecosystem service assessment is critical if urban planners are to ascertain the most comprehensive account possible of the value of urban green spaces for human well-being (Plieninger et al., 2013), thus new approaches to how they are assessed and quantified are essential. The concept of ecosystem services provides a framework for promoting the continued inclusion and addition of urban green spaces in the urban landscape. Quantifying and understanding the transfer of beneficial ecosystem services to human populations is well placed as a means to justify the continued provision of well managed urban green spaces when municipal budgets are increasingly limited.

1.3 The importance of interdisciplinary research in the urban context

Urban areas are defined by their human presence. They can be seen as environments where humans and their associated technologies collide with natural ecosystems (McIntyre et al., 2000); providing habitats for humans and biodiversity alike. Whilst humans are components of the natural ecosystem, much of the technology found in urban areas is not. Urban landscapes provide the backdrop for examining the interactions between the human, technological and natural components of the system with each impacting and affecting the other.

The notion that humans and their surrounding physical environments interact to determine human experiences of the landscape has been prevalent throughout the discourse of defining 'landscapes' (Sauer, 1925; Grossman, 1977; Forman and Godron, 1981; Cosgrove, 1984). One such discussion posits that the landscape aesthetic is a key link between human and ecological processes; being integral in the emotional response of humans to the landscape and resulting uses (Moore and Young, 1978; Gobster et al., 2007).

Despite the connections between human and physical disciplines of geography in understanding how ecosystems affect human populations, ecological studies of urban areas have often lacked acknowledgment of humans in the ecosystems. As a discipline, ecology is defined by investigations of the relationship between ecosystem structure and function (Odum, 1971), played out (usually) in the context of how ecological structure affects ecological functioning, as opposed to the implications of ecosystem structure and functioning on human populations.

Within ecological study, urban areas have come to represent a new biome; however, their definition as 'urban' is based heavily on the presence of humans within them, rather than variables of temperature, precipitation or dominant vegetation typically used to define other biomes (Botkin and Beveridge, 1997). The nature of this definition, and the recognition that urban areas represent the most human dominated of all ecosystems, demonstrates that to understand the socio-ecological processes within these landscapes, it is critical that traditionally separate disciplines intersect and integrate social and ecological investigations (Grimm et al., 2000).

Divergent approaches have emerged to address this from ecologically, sociologically and psychologically anchored viewpoints. The emergence of human ecology in sociology during the 1920s associated with the Chicago School viewed human beings as biological organisms and social beings interacting with their environment (Bubolz and Sontag, 1993). While this was a progressive step in the integration of human and biological study, the focus in the urban context was explaining social changes as a result of urbanisation (Park et al., 1925 cited in Grove and Burch, 1997) rather than the impacts of urban ecosystems directly on individuals within the population. Further recognition of the need for integration resulted in the disciplines of social ecology, environmental sociology and ecosociology which emerged specifically to study interactions between society and the environment (Catton and Dunlap, 1978; Mehta and Ouellet, 1995). These continued the view of humans as integral components of natural ecosystems along with an emergent theme of environmental responsibility for the increasingly detrimental effects of humans on global biodiversity and the interrelatedness of social and environmental problems (Tarman-Ramcheck, 1983; Bookchin, 1988). While these sub-disciplines are successful in integrating social study with ecological concepts, they are inherently social. Given their nature, there is an overwhelming focus on theory rather than scientific investigation in their discussion.

Urban ecology has become the main ecological sub-discipline of investigation of process dynamics in the urban context (Wu, 2008). As an emerging interdisciplinary field in its own right, urban ecology is influenced by both ecological and social sciences (McIntyre et al., 2000). Research aims in this field focus on understanding the human role in urban systems as one of detriment to biodiversity, particularly in human dominated cities (Haase et al., 2014). Given the parent discipline, research in this arena is predominantly undertaken using scientific methods. An increasing theme within urban ecology is the study of ecosystem services in relation to human well-being, a topic of increasing importance within the wider health arena (for example: McLeroy et al., 1988; Grywacz and Fuqua, 2000; Rapport et al., 2003; Boleyn and Honari, 2005; McLaren and Hawe, 2005). Such research is essential in informing urban planning and urban decision making in a variety of context (Pickett et al., 2004). In this way, the field of urban ecology provides another overlap between society and the landscape.

While urban ecology does study the interaction between society and ecosystems, focus has been on one side of this relationship, emphasising the often detrimental impacts of human populations on ecosystems and the debated role of humans within the wider ecological system. Extensive research continues in this theme given the desire to protect and improve biodiversity in an urbanised world. Increasingly, there is a need for urban ecology to be an interdisciplinary field where the social and natural sciences converge (McIntyre et al., 2000), appreciating the reciprocal relationship between the human and natural components of the urban ecosystem. This will enable the expansion of comparatively little investigation into the impact of urban nature on humans, despite the recognition that ecosystems may have significant beneficial effects for urban populations (Maas et al., 2006; Mitchell and Popham, 2007; Lee and Maheswaran, 2011).

The emergence and growth of environmental psychology in recent decades is, to a certain extent, providing such an arena, where the impacts of space on humans are investigated. Naturally such research efforts are focused on implications of space for human behaviour and emotion (Wells et al., 2016). Key to the theoretical framework underpinning research in environmental psychology is the notion that to understand how space impacts the individual most effectively, there must be consideration of the wider social, economic and cultural contexts in which the interaction between individual and space takes place (Saegert and Winkel, 1990).

This thesis includes aspects of all these disciplines, interlinking concepts traditionally seen as independent in order to provide a novel and interdisciplinary approach to investigating urban green space and its interaction, role and influence for the human urban populations. Cultural ecosystem services are perfectly placed as a concept for bridging the gap between different disciplines and research agendas (Milcu et al., 2013). Given their relatedness to human well-being, attitudes and beliefs, cultural ecosystem services highlight powerful linkages between the social and ecological sciences, providing a framework that researchers from numerous disciplines can engage with to gain a holistic understanding of the socio-ecological interactions resulting from human contact with urban ecosystems (Haase et al., 2014).

1.4 The new role of smart devices in urban research

Understanding how the human use of urban green space varies both spatially and temporally is essential in beginning to establish how and when human populations are exposed to the benefits of cultural ecosystem services. Methods of assessing human use of urban green spaces have previously been limited through time and cost restrictions inherent within their methodologies. Thus, new assessment methods are essential to improve understandings of human interaction with ecosystem components of the urban landscape. The rise of smart technologies and social networks offers a solution to this previous limitation in studies of human-ecosystem interaction.

The use of smart devices and phones has seen rapid uptake since their introduction as a consumer product in the 2000s with current ownership in the UK estimated at 76% for adults (Deloitte, 2015) and 90% for 16-24 year olds (Ofcom, 2015). Their use is continuing to grow, penetrating increasingly varied cross sections of society as the technologies become cheaper and more accessible, and their supporting infrastructure more advanced (Ofcom, 2015). The GPS enabled features standard to these devices provide a new way to digitally record individual interactions with urban environments (Frias-Martinez & Frias-Martinez, 2014). In this context, smart devices provide a useful sensor of human behaviour and activity (Raento et al., 2009), particularly in urban areas where their use is commonplace given the desire to communicate instantly and conveniently with others in the fast-paced urban world (Madden et al., 2013).

One particular aspect of smart devices which makes them useful for studying human interaction with urban green spaces is their platform functionality for social networking. The data made available by individuals through social networks about their daily lives, behaviours, (dis)likes and habits means that researchers are no longer reliant on government data as their only source of large scale social data (Murthy et al., 2015). Social networks and the large volume of data they generate about their users, make it possible to obtain information about the complex city system that could traditionally only be obtained through fieldwork (Agryzkov et al., 2016). As such, they provide a source of 'people and environmental sensing' information that can be used to understand factors behind the habits of populations (Silva et al., 2013). This is an example of crowdsourcing which, in its simplest form, refers to a group of people producing data that can be used by third parties to solve a problem (Estelles-Arolas & Gonzalez-Ladron-de-Guevara, 2012).

The vast datasets being produced by such platforms require an increased collaboration between qualitative and quantitative based research methods to enable the most effective data-intensive inquiry across the physical and social sciences (DeLyser and Sui, 2012). Indeed, the growing interest in exploring the social information that these data sources may contain requires methodological innovation that is uniting scholars from both humanities and physical sciences together. It is to this collaborative venture that this thesis contributes, engaging with and critically assessing datasets derived from social networks and apps in the context of urban green space research. In particular, this thesis focuses on data obtained from the social network Twitter, a free microblogging service which enables users to communicate through short statuses and messages of up to 140 characters in length.

1.5 Aims, Objectives and Thesis Outline

The overall aim of this thesis is to investigate the utility of new data sources and methods in understanding how people use and consume urban green spaces, and thus receive the cultural ecosystem services these spaces provide. Explicitly, the aims of this thesis are to:

(i) Demonstrate the utility of data obtained from smart device enabled platforms (social networks and apps) in understanding socio-ecological interactions in urban areas between human populations and urban green spaces and;

(ii) Critically evaluate the use and utility of these data sources for researchers and policy makers.

This thesis has started by providing an overview of the relevant literature, scientific context and theoretical frameworks to the work. Following an introduction of the study area and brief description of the data and methods used, it presents four empirical chapters, all of which have been accepted for publication (Chapter Three [(Roberts, 2017)], Chapter Four [(Roberts et al., 2017)], Chapter Five [(Roberts et al. in press)] and Chapter Six [(Roberts et al. in press)]). Each empirical chapter details the directly relevant literature and methods before presenting data analysis, discussion and critical evaluation. Following these empirical chapters, Chapter Seven presents a discussion exploring the potential of data derived from social networks and apps in facilitating more nuanced understandings of human mobility in an urban context. Finally, Chapters Eight and Nine summarise the findings of this thesis; presenting an overall synthesis and discussion, and concluding remarks respectively.

All the empirical chapters make use of data obtained from the social network Twitter in their investigations and demonstrate its use in a diverse number of applications. Chapter Four also utilises Netatmo (a low-cost consumer weather station) to provide crowdsourced data on meteorological variables, while Chapter Seven uses data obtained from the BetterPoints app alongside Twitter data. Whilst addressing the overall aims of the thesis, each chapter also identifies and addresses more specific objectives of this thesis. To this end, this thesis addresses five specific objectives as follows:

(i) To critically evaluate the use of Twitter data in the assessment of urban green space multifunctionality (Chapter Three);
(ii) To critically evaluate the use of Twitter data in investigating the temporality of physical activity engagement within urban green spaces (Chapter Four);
(iii) To critically evaluate the use of Twitter data in emotional response detection of individuals to urban green space (Chapter Five);
(iv) To assess the most appropriate type of sentiment analysis for Twitter data in the urban green space context (Chapter Six);
(v) To critically evaluate the potential of crowdsourced data from Twitter, BetterPoints and other apps in furthering understandings and investigations of human mobility in urban areas (Chapter Seven).

Following these empirical chapters, a general discussion and concluding comments are presented.

1.6 Justification of this research in Birmingham

This thesis bases its investigations into socio-ecological interactions in the metropolitan area of Birmingham, United Kingdom. Chapter Two provides details of the exact locations of the sample of urban green spaces included for investigation by this thesis; however, it is first useful to consider the relevance of urban green space research in Birmingham.

Green space is a significant component of the cityscape (Figure 1.2) and has become an increasingly prominent feature of the discourse surrounding the city's morphology, reflected in numerous publications such as the 'Green Living Spaces Plan' (BCC, 2013) and the 'Parks and Open Spaces Strategy' (BCC, 2006).

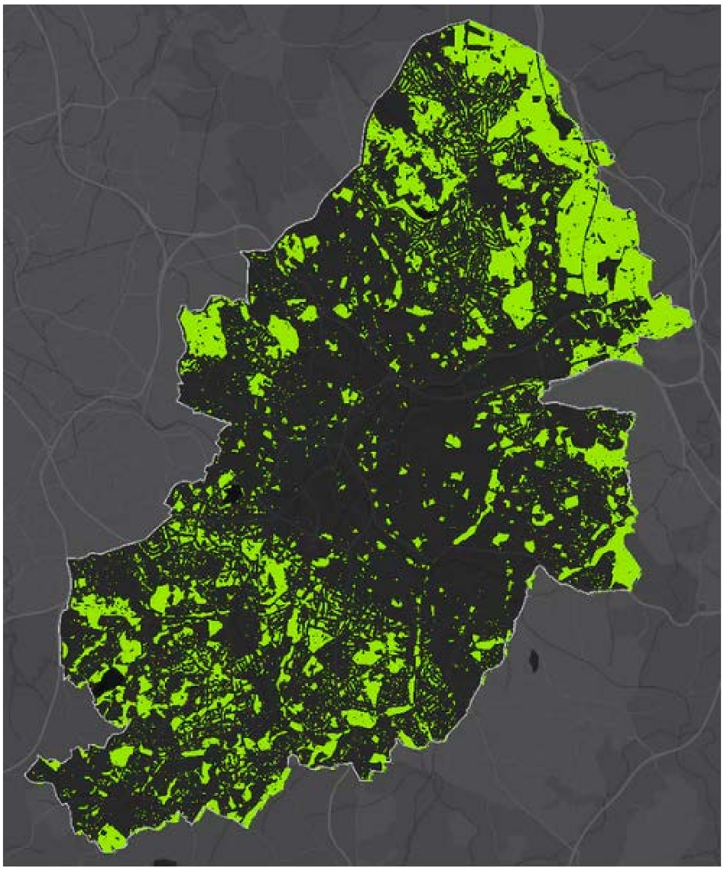


Figure 1.2 *The location of publically accessible urban green space in Birmingham. Taken from the Guardian Online, based on ESRI data (Source: Guardian, 2017).*

Urban green spaces are recognised as having the potential to mitigate and reduce many of the problems encountered by urban dwellers, for example poor air quality, reduced interaction with nature, etc. Birmingham City Council has also specified the need to make more of the physical infrastructure in the city, including green infrastructure, to increase engagement with physical activity (BCC, 2016b).

Predictions from the government-commissioned ‘Healthy weight, Healthy lives’ report (Swanton, 2008) stress that without clear action, 9/10 adults and 2/3 children will be obese by 2050. This presents a significant financial strain on society and negatively impacting the lives of those individuals. Urban green spaces have a significant role to play in facilitating behavioural change and improving both the physical and mental health of populations. Cultural ecosystem services delivered by green spaces include the provision of space for recreation and physical activity, and this service is of increasing importance as a free resource with which individuals can engage with obesity mitigating behaviours (Han et al., 2013; Bedimo-Rung et al., 2005). Given the recent rise in mental health issues in cities (Srivastava, 2009), the mitigating role of urban green spaces as places of relaxation and escape from the city (Soga and Gaston, 2016; van den Berg et al, 2010) mean that encouraged engagement with them could play a significant role in enhancing the mental well-being of urban populations. Economically, it is logical to take efforts to alter trends of obesity and mental health among the population in the present, rather than face an increased health care bill and reduced work force efficiency in the future.

Appreciation of the functionality and utility of urban green spaces for mitigating against the problems experienced by urban dwellers has led to attempts to improve the accessibility, quality and use of green space for city dwellers and increased attempts to value these spaces as natural assets as part of achieving Birmingham’s green vision. The type of green spaces that exist in the Birmingham metropolitan area are extremely diverse and to better understand the different ecosystem services that each provides to the populations that surrounds them, highly spatialized GIS methods and analyses are required. These have thus far remained financially and technically time intensive, limiting their employment in the city. Indeed, much research has posited that high spatial resolutions are the only way in which the localised and highly contextualised problems of a city can be identified and then solutions attempted (Larondelle et al., 2014; Kabisch et al., 2015).

Overall, this thesis pioneers a new source of data which overcomes these financial and time constraints and through demonstrating the use of Twitter data in a range of diverse applications and critically reflecting upon the utility of the data in each context, the foundations of a new cheap and easily accessible methodology are established.

Chapter 2. Methods

This thesis adopts a mixed methods approach to explore the aims set out in Chapter One. Both qualitative and quantitative analyses are undertaken to investigate the versatility of Twitter data and explore the range of information that it can provide. To best identify the range of applications that Twitter data can be utilised in, this thesis does not restrict the research undertaken to one epistemological approach. Indeed, the chapters presented in this thesis have alternate approaches underpinning their investigation.

Chapters Four and Six utilise quantitative methodologies, employing a variety of statistical tests to quantify the relationships between the variables under investigation. These chapters rely on objective, verifiable observations and measurements as the basis for the analyses they undertake; and in doing so maintain the core principles underpinning a positivist approach (Walmsley, 1974). In comparison, Chapters Three and Five utilise the qualitative methodologies of thematic analysis and sentiment analysis, both undertaken in these chapters within a constructivist framework (Guba and Lincoln, 1998); in which the world is understood to be socially and culturally constructed, constantly being experienced and re-experienced by thinking, feeling beings (Bryman, 2001). To best interpret participant observations, received through Tweets, the wider context from which they have been created are of importance in these chapters.

Thus, this thesis does not strictly align itself to one paradigm; rather it employs a research design in which constructivism and objectivism are both employed. It is useful to consider the quantitative information that Twitter data can provide about the human interactions occurring within urban green spaces. However, to work within a solely positivist approach would reduce the experiences and interactions between people and urban green spaces to separate and fragmented happenings and dismiss the insight that can be gained through a constructivist approach. This thesis argues that Twitter data are versatile enough to be employed in the context of both these epistemological approaches, and to demonstrate the utility of these data to researchers, it is necessary that this thesis presents research undertaken within both frameworks.

This methods section introduces the study location and characteristics of the study sites, and also provides an overview of the Twitter, Netatmo and BetterPoints datasets, specifically focusing on their use, size and scope. Please note that the empirical chapters presented herein have been created as standalone papers with the attendant repetition of elements of the methods.

2.1 Study Location

Situated in the English Midlands, Birmingham is the second largest city in the UK with a population of approximately 1.1million people (ONS, 2014); making it the largest city authority outside of London (BCC, 2011). Birmingham experiences a temperate but relatively continental climate due to its inland position (Figure 2.1). The city centre covers an area of 267 km^2, while the wider metropolitan area covers approximately 599 km^2.

The Birmingham metropolitan area is an highly urbanised region and, since the industrial revolution has experienced numerous phases of redevelopment and re-generation. However, the metropolitan area of Birmingham extends further than the urban centres and includes a variety of land uses ranging from high rise urban through to comparatively rural areas. This land use heterogeneity has implications for the amount of green space found in the city. Within the metropolitan area, there are nearly six hundred public parks, open spaces and nature reserves - the most of any European city (BCC, 2016a); making them an extremely significant component of the urban landscape (Figure 2.1). They provide an important resource for the surrounding populations in terms of their contribution to cultural ecosystem service provision.

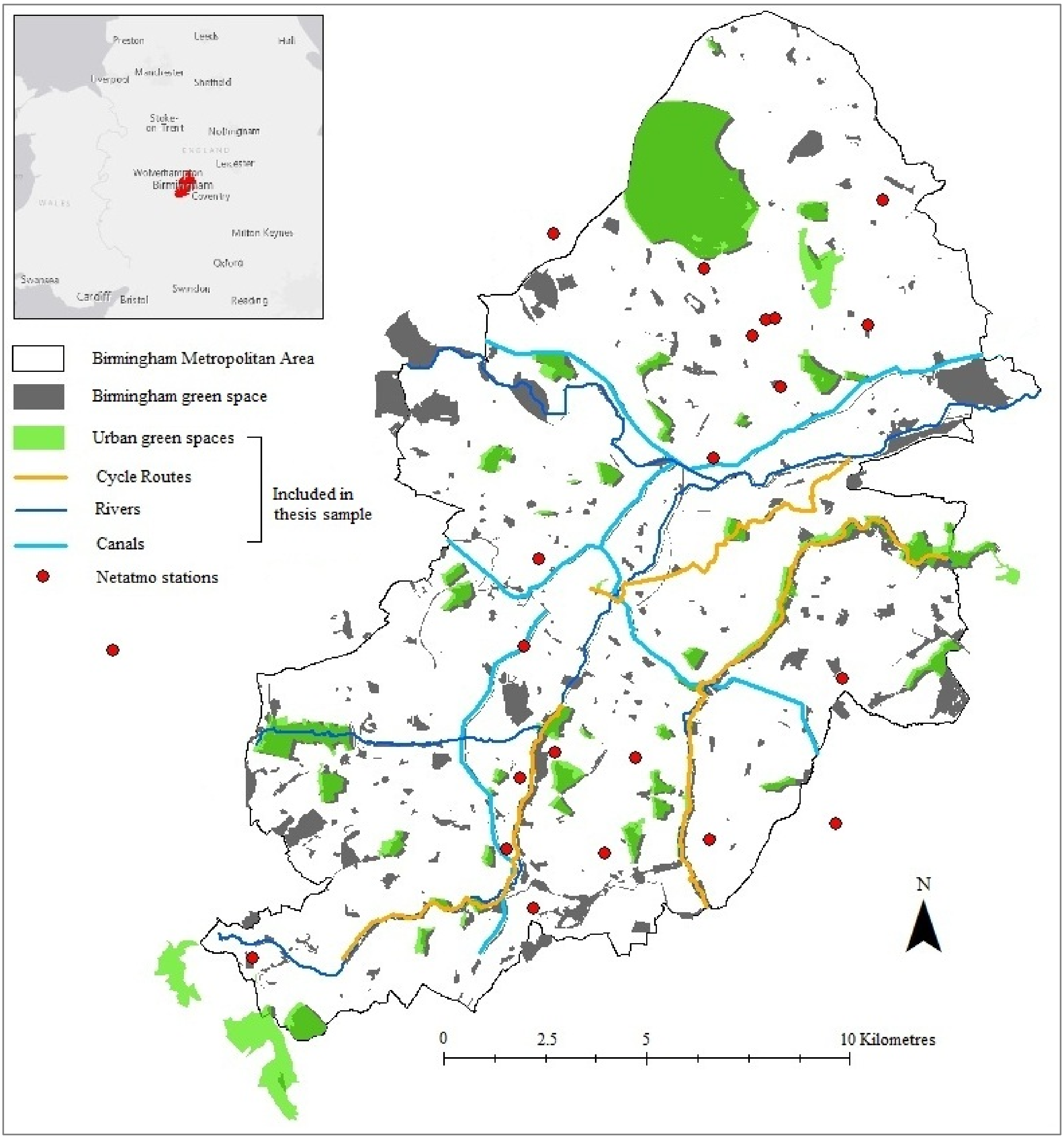


Figure 2.1 *A map of green (and blue) space in the Metropolitan area of Birmingham; and the sample of urban green spaces included in this thesis. Inset shows the location of Birmingham within England.*

A total of sixty urban green spaces in Birmingham are used in this thesis (Figure 2.1). These spaces were chosen to reflect the diversity of urban green spaces found across the city and include parks of a variable sizes and with differing amounts of woodland and water (Table 2.1). They offer a range of services and facilities and are found within different types of neighbourhoods. Alongside the forty six parks, green linear features were also included comprising of the footpaths alongside four rivers and seven canals, and three cycle ways.

Table 2.1 *The diversity of urban green spaces included in the sample in terms of size[1], tree cover, standing water cover and park amenities[2].*

Park Name	Park Area (Hectares)	Categorised Size	% Standing Water Body Cover	% Tree Cover	Number of amenities
Aston Park	22.45	Medium		23.28	10
Billesley Common	29.80	Medium	0	0	6
Birmingham Wildlife Conservation Park	4.03	Small	0	0	4
Brookvale Park	22.40	Medium	35.48	0	7
Burbury Park	4.12	Small	0	0	2
Cannon Hill Park	33.31	Medium	7.49	22.98	12
Cofton Park	61.96	Large	0	5.21	2
Cotteridge Park	10.10	Small	0	15.96	6
Eastside City Park	3.04	Small	1.89	0	2
Edgbaston Reservoir	28.75	Medium	81.12	14.73	6
Fox Hollies Park	15.54	Small	2.55	9.75	4
Handsworth Park	35.45	Medium	5.40	5.85	8
Highbury Park	42.22	Large	0.62	0.96	4
Hillhook reserve	4.62	Small	26.51	72.53	0
Kingfisher Country Park	244.70	Extra Large	2.57	10.96	1
Kings Heath Park	12.52	Small	0.84	0.00	5
Kings Norton Park	9.73	Small	0.00	3.80	4
Kings Norton Reserve	14.35	Small	2.54	10.49	1
Lickey Hills Country Park	198.33	Extra Large	0.13	19.84	0
Manor Farm Park	24.31	Medium	4.15	18.73	6
Mansfield Green	1.40	Small	0	0	0

[1] Categorised size was determined as follows: Small 1-20 hectares, Medium 21-40 hectares, Large 41-80 hectares, Extra Large over 80 hectares.
[2] Park amenities included sports pitches and courts, Multi-use Games Areas, pavilions and club houses, walking routes and boardwalks, fitness trails, green and outdoor gyms, bowling greens, sailing, canoeing and rowing clubs, mini golf courses, skate parks and BMX tracks.

Merecroft Pool	3.44	Small	25.08	0	0
Moseley Bog	18.56	Small	0	60.79	2
Moseley Park	4.89	Small	23.36	0	3
Muntz Park	2.19	Small	0	0	1
New Hall Valley Country Park	97.19	Extra Large	0.31	5.18	3
Perry Park	46.24	Medium	6.81	0	3
Plantsbrook Reserve	10.01	Small	61.99	38.01	0
Pype Hayes Park	42.43	Large	1.43	0	5
Rectory Park	28.95	Medium	0	22.81	3
Rookery Park	6.61	Small	0	0	3
Rubery Cutting	1.02	Small	0	0	0
Selly Park	8.03	Small	0	0	3
Senneleys Park	35.47	Medium	0	4.79	6
Sheldon Country Park	46.23	Large	0	0	7
Small Heath Park	17.16	Small	8.09	15.12	3
Sparkhill Park	10.34	Small	0	1.31	6
Summerfield Park	15.32	Small	0	0	5
Sutton Park	893.48	Extra Large	3.01	30.65	8
Swanhurst Park	17.50	Small	9.17	19.56	2
The Shire Country Park	45.89	Large	0.76	38.29	0
Ward End Park	23.41	Medium	5.24	14.57	5
Waseley Hills Country Park	80.86	Extra Large	0	13.96	6
West Heath Park	15.03	Small	0	2.67	4
Witton Lakes	22.19	Medium	32.76	0	1
Woodgate Valley Country Park	144.20	Extra Large	0	36.63	8

2.2 Twitter

As outlined in Chapter One, the development of technologies, such as smart devices, that support social networks provides researchers with a unique opportunity to access a wealth of information about individuals. The rapid emergence of social networks offers great potential to enable the extraction of human-centred information about a location through technological means, contributing to the concept of augmented space (Cresswell, 2004).

Created and launched in 2006, Twitter is a free microblogging service which enables users to communicate through short statuses and messages of up to 140 characters in length. Any user connected to the internet, on a mobile device or computer, and with a Twitter account has high speed access to the social network. Twitter now reports 313 million monthly active users (Twitter, 2017), with over 500 million tweets uploaded to the network each day (Internet Live Stats, 2017). Users have integrated their engagement with social media sites like Twitter into their daily practices and research from numerous fields is examining this to understand the practices, implications and cultures of these sites, as well as how users engage with them (Ellison, 2007).

Various studies suggest that social media is altering how individuals communicate and socially interact with one another (Kwak et al, 2010; Zhao and Rossen, 2009), with micro-blogging sites like Twitter bringing a new type of communication technology for people to engage with. The posts an individual can create are limited in length, can have multimedia attachments and can reach extensive networks of people in the public domain within a short space of time. Research is beginning to address interesting questions in terms of the social functions of these micro-blogging sites and the information they can provide about engagement with virtual social behaviour. Indeed, Java et al. (2007) have suggested three types of distinct user activities on Twitter: information seeking, information sharing, and social activity. Naaman et al. (2010) go further to describe how much of the information shared on Twitter can be categorised as opinion and 'about me' information, highlighting how Twitter is fast becoming a self-promotion tool for many of its users. Twitter has been found to diverge from the norms of authentic social interaction in that links between users are often not reciprocated (Huberman et al., 2008; Kwak et al., 2010), for example one can follow a user who does not follow them back, creating unidirectional social interactions. This, and the ability to maintain virtual interactions with people who an individual has no contact with in their day to day life are cited reasons as to why people engage with Twitter (Zhao and Rosson, 2009).

There is currently much debate as to how social media enables an individual to present a moderated version of their life and activities, and that the posts a person makes on may not accurately reflect their day to day life, or true emotional state. Indeed, some have posited that social media offers the chance for an individual to alter how they present themselves and create a performance of their lives which differs from reality (Lange, 2007). In addition, users have been found to be selective as to the information they choose to share with others, often choosing information which they feel makes their life seem to be more interesting than they perceive it to be in reality (Marwick and Boyd, 2011). This can be seen as a digital manifestation of how people behave in the real world, with individuals choosing the information they share with (and hide from) others in order to control the impressions other people form of them (Kaplan and Haenlein, 2010). Concerns around user privacy may also have implications for the information that a user may or may not share on a social media platform such as Twitter (Trepte and Reinecke, 2011). The ability of an individual to manipulate the information they present about themselves has implications of using this information as a source of data in research, and while much research (demonstrated in later chapters) has found social media to be an excellent source of information about its users, caveats should remain in place as to the reliability of social media as a source of information in some cases. Whilst social media data offers extensive data to humanistic disciplines and makes social practices and spaces more quantifiable, this data is still subjective and does not necessarily enable the capture of objective truths (Boyd and Crawford, 2012).

Twitter users receive and share information making the social network highly influential in the distribution of information and opinion (Mathioudakis and Koudas, 2010). Its popularity is credited to the ability of users to gain insight into other users without the necessity of having a connection with or to them (Russell, 2013; Suh et al., 2010). Following a person affords a user instant access to another person’s profile, without requiring their permission.

There is no requirement for reciprocal following and the user may not even be aware of who is following them (Weng et al., 2010). Moreover, many pages are open access and do not require any registration on Twitter, unlike other networks such as Facebook. The consistency of Twitter data in terms of the current 140 character limit on tweets means that analysis is more straightforward than other mixed media posts of varying lengths (Highfield and Leaver, 2014).

Twitter data are being increasingly engaged with by the research community in varying contexts because of the advantages they confer over more traditionally acquired data, through means such as participant observation. The free and easily accessible nature of Twitter data through the REST and STREAMING APIs (Application Programming Interface) make it an attractive source of data with which to engage with. Each API makes different tweets available. The STREAMING API enables the streaming of tweets in real-time. In comparison, the REST API provides access to tweets published in the last 7 days and allows queries to be used to search for specific tweets. Both these APIs provide only a random 1% sample of tweets related to the search query. Searches made using the REST API are based on relevance and therefore this source of tweets was considered more appropriate for collating a dataset for use in this thesis. Twitter does also make data available to download through its FIREHOSE service which enables the streaming of all the tweets which match a search query in real-time, rather than the random 1% sample available through the STREAMING API. However, this comes with substantial financial cost making it impractical for use in the majority of academic research and the STREAMING and REST APIs are commonly relied upon to access Twitter data.

Once a connection to either API has been established, data are easy to harvest and require minimal human input for repeated downloads over time. In urban research, the high spatial and temporal resolution that Twitter data provides, affords researchers increased opportunities for multi-scalar investigations, which has been cited as a necessary requirement in the urban context if the many processes on-going within the urban landscape are to be meaningfully investigated in their appropriate context (Larondelle et al., 2014; Kabish et al., 2015). Finally, the geolocation feature, which tags the exact geographic location of a user when the Tweet is posted, is an important parameter enabling an assessment of how people interact with spaces through pinpointing their spatial position at the time they tweet.

2.2.1 Tweet corpus creation and datasets

Two Tweet datasets are used in this thesis: the first is used in Chapters Three, Four, Five and Six and the second in Chapter Seven.

To collate the tweets in the first of these datasets, this thesis used the ‘twitteR’ package which is designed specifically for working with the Twitter REST API and has the required coding functions/libraries already in place. This package makes use of the OAuth protocol, a method which enables third party researchers to access user data without gaining access to passwords and other private information (Hawker, 2010; Russell, 2013), thus ensuring Twitter user confidentiality. The REST API does provide the username of the individual who created a given tweet in the metadata downloaded alongside the tweet text, but in this thesis, no usernames are referenced to ensure complete anonymity of users. Access through OAuth grants a third-party user an access token and an access token secret which act as their credentials to access the user data.

To make a connection through ‘twitteR’ possible, it is first necessary to create an account with Twitter. Using these account details an application account must be created with Twitter which can be done at the following website: https://dev.twissoutter.com/. This process generates a unique API key, API secret, access token and access token secret, which are needed to make a connection to the Twitter API (Figure 2.2).

```
> require(plyr)
> require(twitteR)
> consumer_key <- "……"
> comsumer_secret <- "……"
> setup_twitter_oauth(consumer_key, consumer_secret)
```

Figure 2.2 *The code required to connect to the REST API using OAuth protocol.*

English language tweets were downloaded approximately every ten days from the API to ensure maximal temporal coverage over a period of twelve months, from June 2015 to May 2016. A search query (Figure 2.3) was used to ensure that the tweets downloaded were related to one of the sixty sites included in the study. Pre-processing of the tweet corpus prior to analysis was undertaken to remove any duplicates.

```
> cofton <- searchTwitter("cofton park", n=1000, lang="en")
> coftonnrt <- strip_retweets(cofton)
> cofton.df <- twListToDF(coftonnrt)
> write.csv(cofton.df,file="Cofton Park.csv")
```

Figure 2.3 *Example search query used to download tweets from the REST API.*

The final corpus used in chapters Three, Four, Five and Six includes a total of 10,197 Tweets downloaded during the twelve month period relating to the sixty study sites. Although extensive and sufficient for the analysis undertaken herein, it is by no means a 'big data' set as it relies on the REST API which only provides a random 1% sample of all tweets generated. Moreover, a search query was used to target tweets from pre-determined sample sites; and both these factors reduced the number of available tweets. However, it is robust enough to explore the utility of Twitter data in providing information about how people interact with, think about, value and use urban green spaces.

The dataset used in Chapter Seven draws upon a larger dataset of 56,169 geotagged tweets, provided by Dr Wendy Guan from Harvard University's Center for Geographic Analysis.

2.3 Netatmo

Started in 2014, Netatmo have created commercially available electronic smart devices. Netatmo weather stations (www.netatmo.com) enable the monitoring of atmospheric parameters inside or outside of a building. The devices are commercially available, enabling interested individuals to monitor atmospheric parameters inside or outside a building. As a smart device, their connectivity enables the data they collect to be transferred in near-real-time to a centralised server for storage, where it is then made available for download through the Netatmo API. Netatmo weather stations are part of the 'internet of things' and play a growing role in the provision of crowdsourced data for atmospheric science (Muller et al., 2015; Chapman et al., 2017).

2.3.1 Netatmo dataset

Chapter Four utilises meteorological datasets obtained from Netatmo weather stations. These stations provide temperature (˚C) and rainfall (mm) data. Observations of both these variables are taken by the weather stations at a temporal resolution of five minutes. Eight weather stations within the study area had continuous recordings of the required meteorological variables over the period of study. The closest outdoor Netatmo stations to the study sites were selected for study (Figure 2.4).

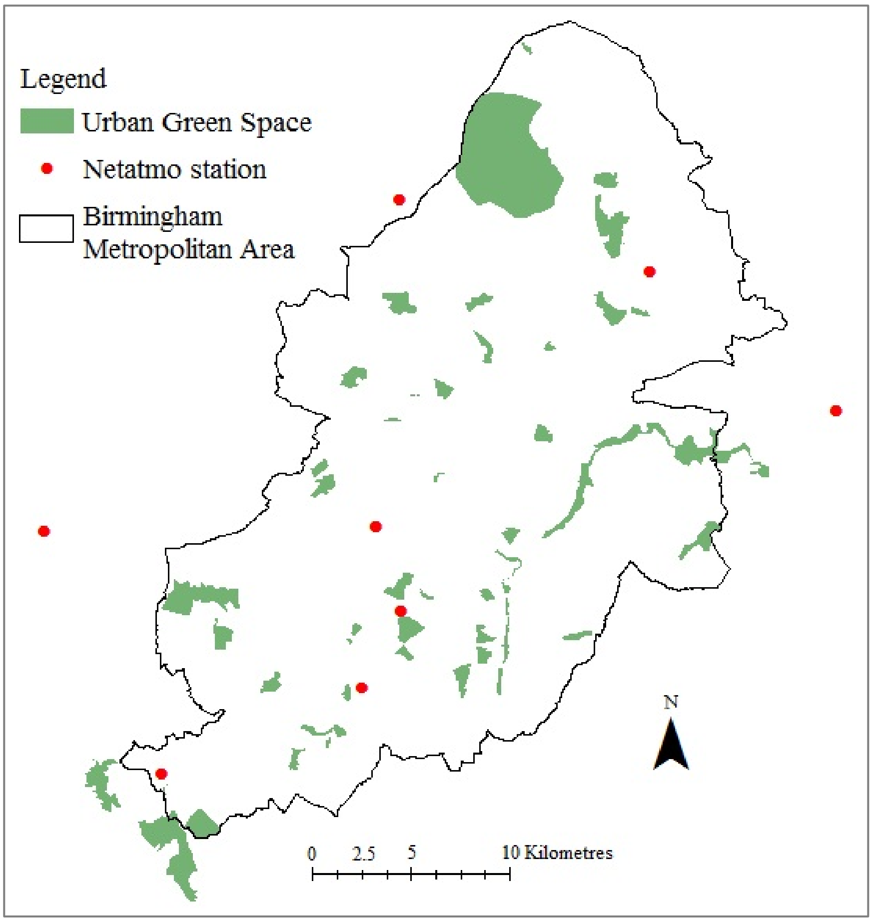


Figure 2.4 *The locations of the 46 urban green spaces in Birmingham included in the study sample and the position of the Netatmo stations used to provide meteorological data.*

The position of these stations was sufficient to capture city-wide sufficient variability in the meteorological variables under study, despite low spatial coverage. The datasets used herein were downloaded via the Netatmo RESTful API.

2.4 BetterPoints

Founded in 2010, BetterPoints Ltd describe themselves as an 'evidence-led sustainability, health and social behaviour change technology company'. Their aim is to provide motivation for individuals to engage with more sustainable and social behaviours. Through their smartphone app interface a user earns points for undertaking a number of activities, including walking, cycling, taking a train or bus and volunteering. The points earned by an individual translate to a financial reward which can either be exchanged for shopping vouchers or donated to a chosen charity. Recognising the power of self-tracking in encouraging positive behavioural change (Choe et al., 2014), the app aims to motivate people to make healthier choices for small rewards as a way to promote long term behavioural change. BetterPoints Ltd are currently active in Hounslow, Hackney, Reading, Sheffield and Birmingham with individuals in these locations being able to sign up to the app for free, log their activities and receive financial rewards.

In the context of this thesis, the data generated through the recordings of users on the BetterPoints app, provides a way of measuring the informal activities that individuals undertake in urban green spaces. These have been found to be an important source of physical activity according to surveys which find that informal activities, such as walking, running and cycling, are more common than formal activity such as team sports (Tzoulas and James, 2010; Greenspace 2007). Utilising the portability, computing and sensing capabilities standard to smart devices, the BetterPoints app provides an innovative data collection tool capable of recording the movement and activity behaviours of individuals through the cityscape.

2.4.1 BetterPoints dataset

Chapter seven makes use of a dataset generated from the BetterPoints app. For the time period over which the dataset used in this thesis was collated, BetterPoints saw an increase in its users from 412 to 878 who range in age from 11 to 70. The dataset is comprised of the recorded activities of these members over a twelve month period from June 2015 to May 2016. During this time users were able to use the app to log their walking, running and cycling activities. All distances and durations of these three activities are included in the dataset used herein. These logged activities are then stored on a central database and available to the company. BetterPoints have kindly given permission for the use of a section of their dataset in this thesis. The dataset comprises of geolocated data points for three activities (running, walking and cycling) arranged in a grid with 10m between each point. Each point has an associated activity count (number of recorded activities which have passed through the point) and user count (number of BetterPoints users who have passed through the point) for each of the four seasons (Table 2.2). The resultant dataset provides information on both user and activity density across the Birmingham metropolitan area. Density data were provided for use in this thesis, as opposed to the individual routes recorded by users to ensure user anonymity/privacy.

Table 2.2 *An example of the BetterPoints data; containing latitude, longitude, activity count and user count information.*

Point ID	Activity	Months	Latitude	Longitude	Activity Count	User Count
1	Walking	DJF	52.4826	-1.9371	741	142
2	Walking	DJF	52.4285	-1.937	715	117
3	Walking	DJF	52.4285	-1.9371	667	104
4	Walking	DJF	52.4826	-1.937	590	98

2.5 Qualitative analysis

In Chapters Three and Five, qualitative analyses are used to explore a range of themes and emotion present in a sample of Tweets. The thematic approach presented in Chapter Three identifies a number of social, political, economic and religious themes present in the Tweet corpus; bringing together themes from a number of literatures, typically disengaged from the cultural ecosystem services debate. In Chapter Five sentiment analysis (Liu, 2012) is undertaken to examine the positive and negative emotional responses of people to urban green spaces, reported in the Tweet corpus. Herein, sentiment analysis describes analysis of the opinions, attitudes and emotions of individuals towards green spaces and their associated entities such as services, organisations and events.

Further details of the approaches undertaken are provided in each chapter.

2.6 Statistical analysis

In Chapters Four and Six of this thesis, a number of statistical tests are used to quantify the difference between datasets as well as to ascertain the significance of relationships between the variables of interest. These include, where appropriate, Wilcoxon signed rank tests, Friedman tests, Student t tests, Welch t tests, Mann Whitney U tests and Spearman's Rho. Fleiss and Cohen Kappa Indexes are also derived to ascertain inter annotator and inter-method reliability.

The specific tests used in each chapter were selected in relation to the research questions and the data structure. Prior to analysis throughout this thesis, data were prepared for analysis using robust exploration protocol (Zuur et al., 2010). Such exploration included checking data for the assumptions of normality, using histograms, QQ plots and Shapiro-Wilk tests; and homogeneity of variance, using Levene tests.

Again, more substantive detail of the statistical analyses employed and justification for test selection is provided in each chapter.

Chapter 3. Using Twitter data in urban green space research: A case study and critical evaluation[3]

3.1 Abstract

Urban green spaces are an important resource for human populations; providing a range of benefits via the provision of ecosystem services. Cultural ecosystem services afforded to human populations by green spaces are of particular relevance to human and societal well-being because of their important roles as spaces of human interaction, economic opportunity, various cultural activities and interactions with nature. To understand how these benefits are received and utilised it is vital to understand when and how human populations interact with urban green spaces. Typically, approaches to investigating when and how human populations interact with urban green spaces have relied on observational and qualitative reporting techniques. Beginning with a critical review of the relevant literature, this chapter highlights a number of limitations associated with methodologies. The development of technologies provides the potential for methodological progression and for these limitations to be addressed and overcome. Using an analysis of 46 urban green spaces in Birmingham as a case study, this chapter investigates the potential of an alternative method centred around Twitter data for examining the interactions of human populations with urban green spaces. A variety of organised events, presenting opportunities for social interaction, economic opportunity and fostering community identity were identified. The utilisation of this technique has the potential to be more cost and time efficient than previous methodologies, as well as enabling

[3] Roberts, H. (2017) Using Twitter data in urban green space research: a case study and critical evaluation, *Applied Geography* **81**: 13-20.

longitudinal study through space and time. However as a new method there are new issues that must be addressed.

3.2 Introduction

3.2.1 Urban green space

Urban green spaces are gaining an increasing amount of attention in both academic and policy arenas. The potential benefits they offer to human populations are increasingly significant in an urbanised society given their potential to facilitate improved human and societal well-being (Chiesura, 2004). The benefits that nature and green space can provide to human populations are broadly termed ecosystem services; encompassing the regulating, supporting, provisioning and cultural benefits that ecosystems provide (Costanza et al., 1997; Daily, 1997; Ehrlich & Ehrlich, 1981; MEA, 2005).

Urban green spaces are beneficial for human populations in a myriad of ways (Keniger et al., 2013) as is clear from the wide range of ecosystem services that have been identified (Costanza et al., 1997; Daily, 1997; Ehrlich & Ehrlich, 1981; MEA, 2005). The benefit of relevance to this case study is the interactions they facilitate between nature and human individuals in the surrounding communities. Provision and engagement with activity in such spaces is thought to facilitate social cohesion (Ewert & Heywood, 1991; Groenewegen et al., 2006), foster social empowerment (Westphal, 2003) and mitigate exclusion and isolation of certain groups (Seeland et al., 2009; Shinew et al., 2004). It has also been found that green spaces have a significant role in developing a sense of place, community identity and ownership of space (Kim & Kaplan, 2004). Urban green spaces are also important in facilitating interactions between people and nature. Such interactions have been found to be beneficial to human well-being; reducing stress and encouraging pro-social and sustainable behaviours (Maller et al., 2006).

While these ecosystem services have been defined, there is a significant lack of understanding as to how these benefits are transferred to and received by urban populations and the circumstances under which this can happen most effectively.

Insight into the interactions between urban populations and green spaces can provide direction for planners to instigate schemes to improve the quality of human life which is imperative given the increasing urbanisation occurring across the globe. Such endeavours to manage urban green spaces most effectively require research founded on an evidence based approach. Similarly, such information can influence local authority decision making and encourage the incorporation of functional and usable green spaces into the urban environment for the benefit of its urban citizens.

Understanding how green spaces are used is a fundamental starting point in improving insights about their significance for human and societal well-being. Therefore, knowing when and how people are engaging with urban green spaces is important, as well as an awareness of the potential barriers that may be preventing their use. Often a space can have multiple, simultaneous functions for different groups of users and understanding these can help in the development of sustainable land use management strategies (Peng et al., 2016).

Previous attempts to examine how and when human populations make use of urban green spaces have followed two methodological approaches. The first of these relies on a qualitative, report based approach in which surveys, interviews and focus groups are used to gain information from participants on a range of topics. These have included the effects of deprivation on green space access (Jones et al., 2009), features of green space which may promote increased use (Schipperijn et al., 2010), personal motivations and barriers to using green spaces (Gidlow & Ellis, 2011), attempts to assess the non-market economic value of green space (Lockwood & Tracey, 1995; Del Saz Salazar & Garcia Menendez, 2007) and assessing the benefits people feel they receive from urban green space in times of heat stress (Lafortezza et al., 2009). Self-reporting techniques have also been employed to assess the impact of park improvements on use and activity (Cohen et al., 2009b). Whilst employed extensively in researching urban green space use, report based methods have a number of limitations. For example, as it is inherently reliant upon participant responses there may be issues with recall and social desirability biases (Evenson et al., 2014) and it is difficult for the information received from participants to be independently validated. Indeed, where validations of questionnaire responses have been undertaken disagreements have been found. For example, Evenson et al. (2013) found only an acceptable agreement (67–82% percent agreement) between the actual and reported park visits of participants, using GPS monitors to validate responses.

The second way in which urban green spaces have been examined utilises an observational approach, treating the urban green space as a study location while researchers record the visitors and activities on-going within it. While used less extensively than report-based methods, such approaches have been employed in a range of contexts related to urban green spaces; including investigations into the types of activities that occur in green spaces (Tzoulas & James, 2010), features of a park associated with physical activity (Kaczynski et al., 2008) and the influence of meteorological variables on green space use (Thorsson et al., 2004). Studies using this method have also investigated the effect of race, age and gender (Cohen et al., 2007; West, 1989) on the use of neighbourhood parks and their significance as a location for physical activity. Specific protocols such as the SOPARC (System for Observing Play and Recreation in Communities) have also been developed in an attempt to produce a standardised approach to observational methods of park use (McKenzie et al., 2006).

SOPARC has been used in investigations of the differences between rural and urban park visits (Shores and West, 2010) and how the installation of fitness zones affects physical activity engagement in parks (Cohen et al., 2010). Observational methods have a significant limitation, requiring multiple observations over different days and seasons to ensure reliability (Cohen et al., 2009a) as park use patterns between specific observation times cannot be reliably estimated. Significant time is therefore required and as a result studies utilising observations tend to lack longitudinal depth.

Consequently, using current methods of observation and subjective reporting the measurement of human interactions with urban green spaces is challenging in terms of achieving consistent results. Methodological progression and new approaches are required to overcome the limitations currently faced (Orr et al., 2014).

3.2.2 Crowdsourcing and social networks data

Various emerging technologies now have the potential to advance assessment techniques of human interaction with urban green spaces: how, when and why people use them, what activities occur within them, and how people feel while using them. Social networks and social media systems enable anyone connected to the internet to provide information about their current location, feelings and activities. As such they provide a source of sensing and information that can be used to understand motivational factors behind the habits of populations (Silva et al., 2013). This is an example of crowdsourcing which, in its simplest form, refers to a group of people producing data that can be used by third parties to solve a problem (Estellés-Arolas & González-Ladrón-de-Guevara, 2012). In the context of this thesis, the crowd is comprised of users of smart technology devices (Kleeman et al., 2008) who share information on social media platforms via the internet. The crowdsourcing of information is becoming increasingly utilised as individuals become progressively connected and accessible in the information age (Brand, 2012).

It is clear that the recent proliferation of mobiles devices are key to crowdsourcing (Kanhere, 2011), especially when obtaining crowdsourced information from social media platforms – a mobile device enables anyone connected to the internet to share their information at any time. These social networks provide a platform to create a human powered participatory sensing network (Demirbas et al., 2010) in which the mobile devices carried by the users become the nodes of the network, connected to provide continuous information to a server. A number of social networks are increasingly present in the day to day lives of millions of people around the world and have already been employed in an academic context (Su et al., 2016).

Created and launched in 2006, Twitter is a free microblogging service which enables users to communicate through short statuses and messages of up to 140 characters in length. Anyone connected to the internet via a smart device or computer, and with a Twitter account has high speed access with the ability to receive and share information. This connectivity along with the large number of users makes Twitter a highly influential player in the distribution of information and opinion (Mathioudakis & Koudas, 2010). Its popularity is credited to the ability of users to gain insight into other users without having a connection with them (Russell, 2013; Suh et al., 2010). Following a person on Twitter affords a user instant access to another person's profile without the need for the other to give permission, follow them back, or even be aware of them (Weng et al., 2010). Indeed many pages have open access status and do not require any sign up to Twitter, unlike other networks such as Facebook.

Information obtained from Twitter has already been used successfully in urban research. Tweet information has aided land use classifications of urban areas (Noulas et al., 2011; Frias-Martinez & Frias-Martinez, 2014; Zhan et al., 2014) and has been used to investigate the emotional responses of people to urban spaces (Bifet & Frank, 2010; Hauthal & Burghardt, 2013; Klettner et al., 2013). It has also been shown to be useful in following how information spreads through urban areas (Malleson & Andresen, 2015; Yardi & Boyd, 2010) and extension apps can be used to monitor a range of environmental variables (Demirbas et al., 2010).

This chapter draws on the successes of such studies and sees them as justification for the inclusion of Twitter data in urban research. While Twitter data have been used to investigate cityscapes in general, it has not yet been applied to the study of urban green spaces despite their significance as components of the urban landscape. This chapter provides a first introduction of the utility of Twitter data in the study of urban green space and the potential of crowdsourced information in improving understandings of such spaces and their importance for human populations. Twitter has been selected over other social networks due to the ease of accessing public data as well as the large numbers of users on the network generating this information.

3.3 Methodology, dataset and thematic analysis

The method described herein explores the potential of Twitter data as a source of information about human interactions with urban green spaces. To gain access to this publicly available data, it is necessary to connect to the Twitter API. R Studio was used as the interface through which connection to the Twitter API was made. The ‘twitteR’ package is designed specifically for working with the Twitter API and the necessary coding functions are already in place. Crucially this method makes use of the OAuth protocol, a method which enables third party researchers to access user data without gaining access to their password and other private information (Hawker, 2010; Russell, 2013). Access through OAuth grants a third party user an access token and an access token secret which act as their credentials to access the user data. Using this package a range of metadata is returned alongside the tweet text, as shown in Figure 3.1.

	A
1	text
2	Loads of FREE #ParkLives fun at Aston Park (Fri 5pm) http://t.co/jWYCXmUbXy http://t.co/VdxZ1P7Iz4
3	Aston Hall and Park #birmingham #8000acres photo courtesy of http://t.co/vtUIYEdfMr http://t.co/bzMBMgkvx2
4	Another of Aston Park #birmingham #parklife #B6 Photo courtesy of http://t.co/DH6rne2MNJ http://t.co/LMf3RhaqFB
5	Great football, great drama , great sportmanship , great day at the Aston Park Rangers Football Tournament today http://t.co/4gfcUayGYx

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource	screenName	retweetCount	isRetweet	retweeted	longitude	latitude
FALSE	0	NA	19/05/2015 10:22	FALSE	NA	6.01E+17	NA	<a href="http:	[illegible]	0	FALSE	FALSE	NA	NA
FALSE	7	NA	19/05/2015 09:13	FALSE	NA	6.01E+17	NA	<a href="http:	[illegible]	10	FALSE	FALSE	-1.90145	52.452
FALSE	3	NA	18/05/2015 19:46	FALSE	NA	6E+17	NA	<a href="http:	[illegible]	0	FALSE	FALSE	NA	NA
FALSE	1	NA	18/05/2015 17:16	FALSE	NA	6E+17	NA	<a href="http:	[illegible]	2	FALSE	FALSE	NA	NA

Figure 3.1 *Example .csv file containing text and metadata information returned from the Twitter API.*

To obtain the tweets for study, a search was made of Twitter's REST API using the park name as the search query (e.g. "Aston Park"). This was the only search term used in order to obtain the full range of tweets related to each park and prevent restriction or biasing of the tweet responses to certain types of activity. Tweets were then manually screened to ensure those included in the sample were relevant and reported an interaction with the specified urban green space.

A basic example is now presented demonstrating the utility of such Twitter data in the assessment of human use of urban green space. The study sites are located in Birmingham, the second largest city in the United Kingdom with an estimated population of 1.1 million (ONS, 2014). Within the metropolitan area, there are nearly 600 parks, public open spaces and nature reserves (BCC, 2016a), the most of any European city. They provide an important resource for the surrounding populations in terms of their contribution to cultural ecosystem service provision.

Using Twitter data collected from tweets concerning urban green spaces in Birmingham, the range of information that can be obtained using crowdsourcing and social network data is illustrated. A case in point is the variety of organised activities found to occur in these green spaces and their importance for social interactions, economic opportunity and community identity is subsequently discussed. A thematic approach was taken in the subsequent analysis bringing together themes from a number of literature typically disengaged from the ecosystem services debate, including economics, social policy and cultural studies.

3.4 Results

Taking the summer months of 2015 (June, July August) as the study period, 24 out of 46 urban parks and green spaces were identified as hosting one or more organised event/s. The locations of these parks are given in Figure 3.2. Parks were chosen to create a sample that reflects the variety of parks in Birmingham. Parks of varying characteristics were selected based on their size, the presence of woodland and water bodies, the presence of a number of different amenities, and their status as Green Flag parks, Nature Reserves and Active Parks locations.

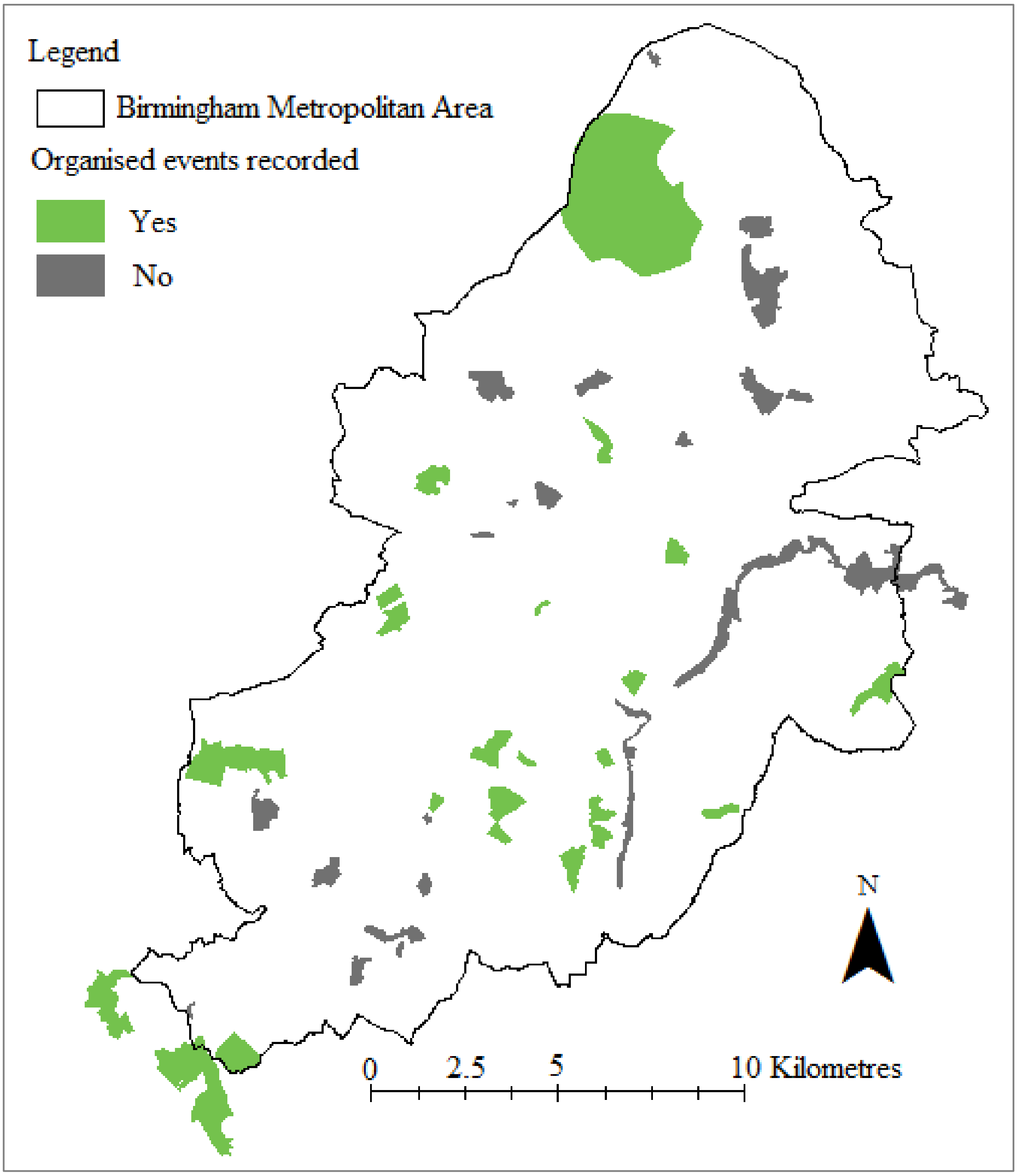
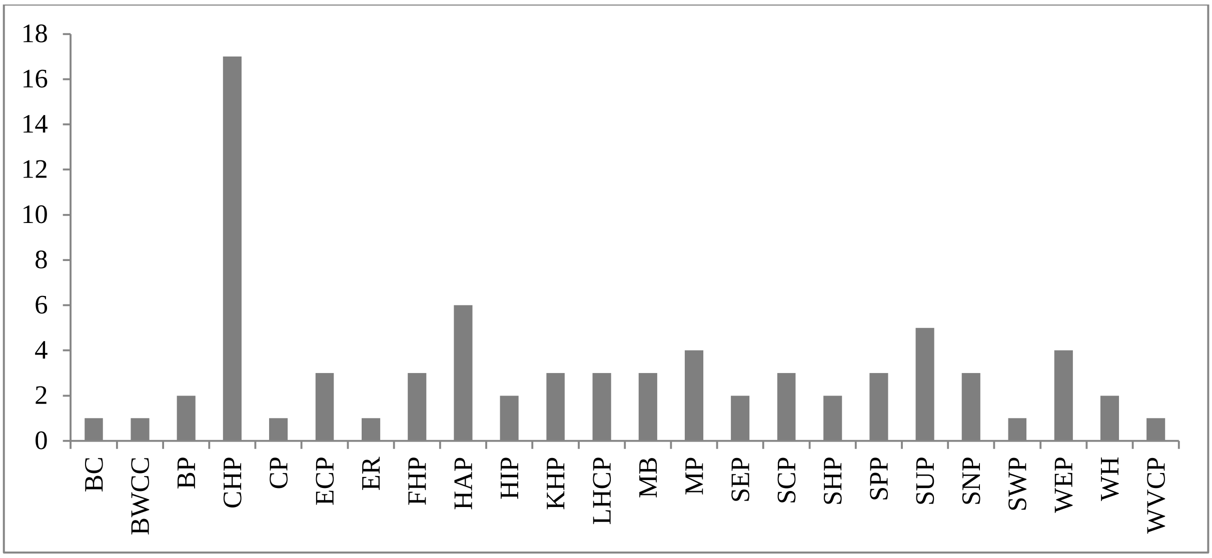


Figure 3.2 *The locations of parks in Birmingham hosting one or more organised event(s) during the summer months of 2015 where green indicates the occurrence of organised events and black indicates the remainder of the sampled parks.*

From a total of 2847 tweets received over the study period 793 tweets relating to 61 separate events were identified, shown in Figure 3.3. Tweets were categorised manually based on an assessment of their text and image content into 11 categories, one of which encompassed organised events. Other categories included physical activity, non-physical activity, nature related activity, charitable activity, economic activity, volunteering, political and religious focused tweets and information based tweets. The organised events category has been utilised herein to ensure a robust number of tweets for analysis.

__Figure 3.3__ A word cloud showing the events captured using Twitter. Word size relates to the number of tweets received for each event, with larger words signifying more received tweets.

The number of events recorded in each park ranged from 1 to 17 (Figure 3.4). The highest number of tweets relating to a single event occurred at the Fusion Music Festival (Cofton Park) with a total of 335 tweets recorded. This is unsurprising given the size and popularity of the event with both locals and those from further afield who travel to the event, as was made clear in the tweets received.



BC Billesley Common **BWCC** Birmingham Wildlife Conservation Centre **BP** Brookvale Park
CHP Cannon Hill Park **CP** Cofton Park **ECP** Eastside City Park **ER** Edgbaston Reservoir
FHP Fox Hollies Park **HAP** Handsworth Park **HIP** Highbury Park **KHP** Kings Heath Park
LHCP Lickey Hills Country Park **MB** Moseley Bog **MP** Moseley Park **SEP** Selly Park
SCP Sheldon Country Park **SHP** Small Heath Park **SPP** Sparkhill Park **SUP** Summerfield Park
SNP Sutton Park **SWP** Swanhurst Park **WEP** Ward End Park **WH** West Heath Park
WVCP Woodgate Valley Country Park

Figure 3.4 *The total number of events recorded per park, obtained from captured tweet information.*

The events identified provide opportunities for individuals to engage at a range of scales, from global events such as the Rugby World Cup (Eastside City Park), national events such as World Food Day (Cannon Hill Park) and the Summer Solstice (Lickey Hills), regional events such as the Big Hoot and Eid celebrations (Small Heath Park) and finally local events such as Edgbaston Regatta (Edgbaston Reservoir) and Acocks Green Carnival (Fox Hollies).

The overwhelming majority (80%) of the events identified were local events (Table 3.1), reflecting the important role of urban green spaces in providing a location where members of the local community can come together and socialise (Low et al., 2009). Previous research has identified the role that urban green spaces play in developing an individual's sense of identity and feeling of connectedness with others (Kim & Kaplan, 2004). Such events may help to achieve this in the local communities in Birmingham. They may also be particularly important for older adults with limited mobility as being able to meet people in their local area is important for them to maintain social ties and a sense of connectedness to the community (Kweon et al., 1998).

Table 3.1 *The scale of the events identified in the sample tweets.*

Scale of event	No. of events of each scale
Global	3
National	3
Regional/City wide	6
Local	49

From the tweets received it is also possible to identify the role of urban green spaces as places of engagement with a range of social, political and religious ideologies. For example, the Eid festival (Small Heath Park), Refugee week events, Vegan Picnic and Pankhurst Picnic (Cannon Hill Park) bring likeminded people together through a shared faith, or perspective. This again links to the importance of urban green spaces in the development of individual and community identity. In agreement with Mitchell (1995), urban green space is shown to be an important space into which different religious, social and political perspectives are brought and celebrated. Other cultural events were shown to occur, with music events such as the Birmingham Mela (Cannon Hill Park) providing the chance for communities to participate in cultural activities including music, dance and art.

The urban green spaces sampled also provided space for a range of activities aimed at facilitating social inclusion and empowerment for groups facing social isolation or other difficulties. Youth projects such as the Girls Youth Hub (Cannon Hill Park) and Sparkhill Youth Project (Sparkhill Park) have previously been identified as having an important role in social inclusion and integration of young people from a range of cultural backgrounds (Seeland et al., 2009). The opportunity that some events create for an individual to meet with others in a similar position to their own can also be beneficial. For example, Brummy Mummy Meetups enable new mothers to meet, socialise and discuss issues they may be facing.

The facilitation of interactions between humans and nature is an important role provided by urban green spaces (Maller et al., 2006). More natural environments have been found to have a restorative effect on cognition, providing an environment with less stressor and a variety of intriguing stimuli (Berman et al., 2008; Kaplan, 2001). Events which focus on bringing people into contact with nature such as Bioblitz, the Big Bog Lunch and flower planting help to facilitate these interactions and bring about improved mental well-being.

While the economic potential of urban green space has been considered extensively from the perspective of the whole city in terms of increasing land value, storm water management and non-market assets (Smardon, 1988; Del Saz Salazar & Garcia Menendez, 2007; Millward & Sabir, 2011), few studies have accounted for their role as discrete spaces for economic activity to take place. 40 of the events identified (66%) had an economic element, providing local businesses, charities or larger organisations with the chance to increase brand exposure and make financial gains through participating in them. Food festivals and summer fetes are a particular example where local businesses set up stalls and are provided with an opportunity for engagement with the local community. On a larger scale, music festivals such as Fusion Festival (Cofton Park) exemplify the opportunities provided to a range of sectors from entertainment, food, security and logistics. Charity events such as those taking place in association with Refugee Week (Cannon Hill Park) provide fundraising opportunities for charities as well as promoting pro-social behaviour and improving individual well-being (Thoits & Hewitt, 2001).

From the three month study period it was possible to elicit a large amount of information from Twitter as to the events occurring in the sampled urban green spaces. The implications these have for individual and societal well-being are discussed and bring together themes from a number of disciplines.

3.5 Critique

This discussion has explored the potential of using crowdsourced data from Twitter in explorations of human interactions with urban green spaces. This method was proposed following the identification of a number of limitations with the previous observational and self-report based methodologies employed; and the successful use of Twitter data in a range of urban related research. It is important to evaluate this method and highlight its utility compared to previously utilised methods, as well as identify any limitations which have become apparent in this use of Twitter as a data source to inform human use of urban green spaces.

3.5.1. Advantages of data collection using Twitter

A large criticism of observational approaches is that to be reliable, they require extensive repeat measurements at the same location (Cohen et al., 2009a), incurring large time and cost expenditures. Twitter data do not have this limitation; indeed tweets can be captured with ease as frequently as necessary providing an opportunity for more measurements to be taken at no extra time or financial cost, achieving greater longitudinal depth as a result. Compared to self-report based methods, tweets are often posted with a photograph giving visual evidence for validation purposes.

The method also provides an unobtrusive method of non-participation which is easy to replicate, improving the potential for a standardised approach to be developed. Being free, publicly available and instantly accessible, the data collection method using Twitter incurs no financial cost and takes significantly less time compared to previously used methods. The consistency of Twitter data in terms of the 140 character limit on tweets means that analysis is more straightforward than other mixed media posts of varying lengths (Highfield & Leaver, 2014). These attributes mean Twitter data are well placed to provide information on the well-being, behaviours and activities occurring within communities (Nguyen et al., 2016).

3.5.2 Issues identified with using Twitter

Despite these benefits, the use of this method also raises some issues which should be taken into consideration in order for it to be utilised most effectively. The first of these is that crowdsourcing information via social networks limits the base population to which one is investigating and there is a need to discuss the inherent biases in these datasets (Hannay & Baatard, 2011). Crucially, those members of urban populations who do not own a smart device are excluded from the sample population. This can have implications for examining the use of space by various sectors of the population such as older people (aged 75+) who show disproportionate levels of non-engagement with these forms of technology (Zickuhr & Madden, 2012). Various spaces throughout the cityscape are supposed to be spaces where all members of the community can come together, but the users of social media do not reflect this diversity (Schwartz & Hochman, 2014) and are therefore not a truly representative cross section of the population. It should also be noted that very limited demographic information is known about the population from which tweets are received. Information about age, occupation or ethnicity is not available through the method described herein which may limit the type of investigation which can be carried out using this method. The metadata provided with the downloaded tweets can go some way to addressing this problem. For example it is possible to ascertain the gender of Twitter users through a search of their profile name on Twitter.

Numerous studies have been undertaken to try to determine the types of people who engage actively with social media (Bendler et al., 2014; Coleman et al., 2009) as a means to assess source credibility. As a general rule, extroverts tend to be more frequent users of social media (Correa et al., 2010) with adults (aged 18–49) making up an increasingly large proportion of those actively engaging through posts (Lenhart et al., 2010). Subsections of the population may be missing in the received dataset due to the inherent biases in using this type of technology; however because no demographic information is available it is difficult to know the direction in which the sample is non-representative.

Methods based on crowdsourcing make use of mobile devices through which people communicate and create a network. This creates issues associated not with the data collection method itself but those that need to be taken into account to understand what limitations there may be on the data available for capture by the method. The quality of internet connection can vary substantially between mobile networks and signal may be intermittent in some areas. An area with limited or no connection to the internet may lead to areas with no recorded use which may not necessarily be an accurate reflection of reality (Chatzimilioudis et al., 2012). While this can limit the production of data, in urban areas such as Birmingham, poor internet connection and mobile phone coverage are unlikely to be problematic as much of the city has 4G coverage. Appropriate selection of where this method is employed can overcome this obstacle to effective use.

User privacy and the ethics of obtaining data in the ways described herein is an important area of consideration when engaging with crowdsourcing through social media (Ma et al., 2008; Burghardt et al., 2009; Vicente et al., 2011). Being able to access the necessary information without compromising the privacy of the user is extremely important to users and researchers alike. This is not a significant issue using the method described in this thesis as only public Twitter accounts are used to provide information, i.e. those who have enabled anyone to view their profile. The use of the OAuth process also addresses the need for privacy and data protection ensuring no personal account details are accessible.

3.5.3 Improving the robustness of a Twitter captured dataset

It should also be noted that the data gleaned from social networks are rarely produced with the aim of it being utilised in scientific research. There may be inaccuracies in their narrative that seem inconsequential to the user but may have significant implications for the research output if utilised by the researcher (Flanagin & Metzger, 2008). While this chapter has shown the potential of Twitter in generating a dataset suitable for investigating human interaction with urban green spaces, there are a number of ways in which the robustness of such a dataset can be improved for research.

Improvements could be made to the resultant dataset by actively engaging and encouraging people to tweet about a specific subject or location of interest. A tried and tested way in which this is achieved is to create a hashtag unique to the study which individuals could be encouraged to add to their tweets. This hashtag could then be inputted into the search query to pull out relevant tweets. This is already being utilised by political campaigns and commercial companies to enable the tracking of Twitter responses to their product or ideas and the creation of a cyber-community who interact together through the use of specific hashtags. While not a traditional approach to data collection, this inductive approach could be employed in the research community with project specific hashtags or accounts affording new opportunities and the creation of a more robust dataset. Success has already been seen to this end with the use of the @ecorecordings account encouraging citizen science engagement with nature sightings.

With respect to the method described in this thesis which connects to the Twitter API, it is important to note that an exhaustive source of tweets is not returned. Any search made to the API provides information of tweets produced in the last 10 days or so. One way to overcome this limitation is to make use of the Firehose API, a feed provided by Twitter that allows access to all public tweets. A significant problem to the use of the Firehose data however is the restrictive cost, as well as the amount of resources required to retain the Firehose data (servers, network availability, and disk space). To ensure maximum possible capture of tweets when using the Twitter API it is advisable to make regular searches to the API approximately every 10 days improving the completeness of the dataset received. Examination of the metadata downloaded with the tweet text, shown in Figure 3.1 can help to provide information about the tweets received. Information such as the time and date of creation and name of the creator provides context to the dataset and improves robustness.

3.6 Conclusion

This chapter has presented a method to investigate human use and interaction with urban green spaces following a critique of the current approaches employed to this end. Twitter is presented as a source of data which can be gathered through crowdsourcing. An example case study of 46 locations in Birmingham, UK has shown the potential of this new approach over a three month period. Twitter data were found to be successful in providing information about the range of organised events occurring in these urban green spaces, indicating the diversity in how urban populations make use of them. A high prevalence of local events was identified along with the provision of opportunity for engagement with regional and national events. The study sites were found to host a range of activities facilitating community engagement with social, cultural, political, religious and nature based events, while also providing space for a range of economic activities.

In comparison to observational and subjective reporting methods which have been used previously in this area of research, the method presented herein offers a number of benefits. These include the free, publicly available and immediately accessible nature of Twitter data, improved longitudinal depth that this method affords and potential to produce a standardised procedure to investigations. It also addresses the time and cost constraints identified with previous methods.

While this method has been demonstrated successfully herein, there is a need to identify a number of confines which much be addressed in order to utilise it most appropriately and effectively. These include privacy issues, biases in the received datasets and a lack of demographic information about the individuals included in the dataset.

Chapter 4. Using Twitter to investigate seasonal variation in physical activity in urban green space [4]

Preface

Chapter Three of this thesis has demonstrated that Twitter data can be used to obtain information on the broad array of social, economic, religious and political activities which occur in urban green spaces. This provides text-based evidence for the significant role these spaces have in providing cultural ecosystem services to urban populations, and establishes Twitter as a useful source of data in identifying the cultural ecosystem services provided by urban green spaces.

Given that evidence-based decision making is common practice in urban planning (Sallis et al., 2016), a logical progression is to investigate how useful Twitter data could be in informing analytical investigations into cultural ecosystem service provision and to ascertain if it is a viable source of data for statistical analyses. Whilst text-based identification of cultural ecosystem service provision is a useful first step, there is oftentimes a requirement for numeric data from which significant relationships or differences between groups can be confirmed.

Physical activity engagement is one such cultural ecosystem service where numeric data is of critical importance and objectively monitoring the numbers of people who engage with different types of outdoor physical activity is essential for providing baseline figures against which future measurements can be compared (Frank et al., 2005). Given the need to reduce obesity levels among children and adults in the West Midlands from 24.4% (BCC, 2014) and 27% (HSCIC, 2016) respectively, urban green spaces within the urban landscape represent an

[4] Roberts, H., Sadler, J. and Chapman, L. (2017) Using Twitter to investigate seasonal variation in physical activity in urban green space, *GEO: Geography and Environment* **4 (2)**: e00041

important resource where individuals can be encouraged to engage with both independent and organised physical activities. Thus, the provision of space for physical activity engagement is an important cultural ecosystem service recognised by urban planners (Bedimo-Rung et al., 2005; Sallis et al., 1998). In order to monitor changes in physical activity engagement successfully, objective data with high temporal resolution is required. Chapter Four examines whether Twitter represents a viable source of such data and in doing so addresses the following objective: to critically evaluate the use of Twitter data in investigating the temporality of physical activity engagement within urban green spaces.

4.1 Abstract

To understand how the benefits of outdoor physical activity in urban green spaces are transferred to human populations, consideration must be given to when people are using them, what they are using them for and what factors may affect the use of space. This chapter critically evaluates the use of crowdsourced Twitter data in an assessment of physical activity engagement in urban green spaces in an attempt to investigate the potential of this data in investigating urban socio-ecological interactions. A case study is presented in which Twitter data are used to assess the variance of physical activity engagement between two seasons (summer and winter). A number of factors including meteorology, park characteristics and amenities and the role of organised sports events are explored in order to explain the observed findings. Understanding how physical activity engagement in urban green space varies seasonally is important in ensuring policy interventions to increase physical activity are targeted most effectively.

4.2 Introduction

Approaches to public health have become increasingly interdisciplinary in attempts to account for socio-ecological interactions (Collins et al., 2000; Forget and Lebel, 2001, Kabisch et al, 2015), with increasing consideration given to the potential salutogenic impacts of the surrounding physical environment on human health (Tzoulas et al., 2007). Ecosystem approaches to health systematically recognise that environmental systems support human health and well-being, and seek to place human populations at the centre of considerations about development and ecosystem management. Socio-ecological approaches look to improve human health through the management of natural ecosystems alongside more direct forms of interventions into human behaviour. However, isolating the links between human well-being and ecosystem processes is difficult because of the complexity of interactions between ecological and social systems (Kittinger et al., 2009). Moreover, attempts to investigate ecological processes and human well-being derive from disparate spheres. Accordingly, there is a need for greater integration of themes across environmental and health research (Haines-Young et al., 2010; Liu et al., 2007) to better understand the complex, multi-scale mechanisms underlying the correlative relationships observed (Scholes et al., 2013).

Green space has been shown to have numerous beneficial effects for human populations, through approaches loosely assembled under the umbrella of ecosystem services (Costanza et al., 1997; MEA, 2005). This chapter focuses on the important role that urban green spaces provide as locations for a range of outdoor physical activities within cities. Participation in physical activity is seen as increasingly important, given the levels of obesity and weight related poor health conditions in the United Kingdom. In England, 62% of adults are overweight or obese (HSCIC, 2015) putting them at increased risk of diseases such as diabetes, cancer, heart disease, stroke and liver disease. Alongside other lifestyle behaviours such as a balanced diet, not smoking and not drinking to excess, physical activity and exercise are a means of reducing obesity and therefore risk to these diseases (GOS, 2007). Physical activity is therefore being is promoted as a public health priority and numerous interventions have been implemented to encourage active participation of people in high risk groups. Parks have been shown to have a significant role in supporting the physical activity of local populations (Han et al., 2013). Indeed, the role of parks and urban green spaces as a location for activity interventions and as a resource to satisfy current physical activity requirements has, for a while, been recognised by decision makers (Bedimo-Rung et al., 2005; Sallis et al., 1998). The effectiveness of these schemes, however, is dependent on engagement with physical activity, which in itself has been found to correlate with a number of factors including accessibility of space (Sallis et al., 1990), park facilities (Krenichyn, 2006), neighbourhood aesthetics (Hoehner et al., 2005), traffic (Tropod et al., 2001) social support from friends and family (Brownson et al., 2001) and perceived neighbourhood safety (Suminski et al., 2005).

The public health literature has identified parks and urban green spaces as common places for physical activity for urban populations (Bedimo-Rung et al., 2005; Cohen et al., 2007; Maas et al., 2008; McCormack et al., 2010). Indeed, studies focused on specific sub-groups have found them to be a significant resource for physical activity for adolescent girls (Cohen et al., 2006), the elderly (Tinsley et al., 2002) children (Muñoz, 2009) and those living in low income households (Lee et al., 2005). Such studies have been useful in emphasising the role of urban green spaces in providing space and opportunity for engagement with physical activity; however, they typically lack specificity on the relationship between park locations and their use by urban populations (Kaczynski et al., 2008). More recently attention has been focused on the effect of park size, park features and distance from residential area on physical activity levels, however, the effects of seasonality and weather conditions are still largely overlooked as a potential determinant of outdoor physical activity (Ergler et al., 2016; Humpel et al., 2002; Tucker and Gilliland, 2007).

In the United Kingdom, the substantial variation in weather within and between seasons may have a significant effect on the use of parks as spaces for outdoor physical activity. Adverse weather conditions have been previously identified as a personal barrier to engagement with a variety of outdoor activities (Lee and Maheswaran, 2010). Rainfall, cold temperatures and icy conditions have all been identified as meteorological barriers across a range of social groups, with particular effect on the engagement of physical activity in the elderly and young children (Belza et al., 2004; Edwards et al., 2015). Despite the progress into understanding the associations between meteorological conditions and physical activity that these studies have provided, there has been recognition that attempts to investigate these relationships have lacked objective assessment (Chan et al., 2006), relying too heavily on self-report methods of data capture. Substantial variation in daylight hours between seasons (ranging from 8 hours in winter to 16 hours in summer) is also an important factor in explaining physical activity variation (Beighle et al., 2008). However, it has been given relatively less attention than other influential factors such as age (Floyd et al., 2011), gender (Kaczynski et al., 2009) and ethnicity (Gordon-Larsen et al., 2000). It is essential for policy makers to understand the seasonal variation in physical activity, and the mechanisms behind this variation, to ensure that interventions aimed at increasing physical activity can be implemented most effectively (Beighle et al., 2008).

Geographical approaches are well placed to investigate the relations between seasonality and social practices and studying cultural practices may provide insight into the varying effect of the seasons on different social groups; with the potential for this information to generate novel positive interventions for enhancing outdoor experience (Hitchings, 2010). Previous methodologies investigating seasonal variation in engagement with outdoor physical activity have followed two main approaches: observational and subjective reporting. While these approaches have their merits, there are significant shortcomings to both. Observational methods, such as counting individuals (Joseph and Maddock, 2016; Suminski et al., 2008), are time-consuming and lack longitudinal depth. Subjective reporting methods, such as surveys and self-report questionnaires (Salmon et al., 2003), suffer similar time and cost constraints as well as other limitations such as participant recall bias (Sallis and Saelens, 2000). The methodological challenges of both these traditional approaches mean that innovative methods are needed if research is to engage more actively with the effect of seasonality on physical activity behaviours.

This chapter presents one such method, using crowdsourced data from the social network Twitter, to provide information about individual engagement and use of urban green spaces for physical activity. The benefits of this approach to the researcher are extensive including a reduction in both financial and time expenditure, given that the data are freely available to download at as frequent a time interval necessary. Given that financial and time restraints are cited as reasons why studying seasonal variation in outdoor behaviours has been neglected (Ergler et al., 2016), the opportunities afforded by this approach are considerable. Access to social network data also provides a larger sample size than is feasibly obtained through observation and survey approaches and may also provide more heterogeneous data than highly structured methods due to the diversity of social network users (Elwood, 2008).

This chapter investigates the seasonal differences in physical activity participation in urban green spaces reported by individuals on Twitter, and the impact of a number of variables on how and when urban populations on Twitter report using urban green spaces as a location for outdoor physical activity. The influence of meteorological variables (temperature and rainfall), hours of darkness, weekday, organised sports events and park characteristics are considered. In doing so the ways in which people use urban green space, and how this varies over a range of temporal scales is examined. This new methodological approach to capturing physical activity engagement within urban green space is presented in an attempt to overcome limitations of previous research. Rather than being reliant upon the results of subjective reporting and observational data, this new approach demonstrates the potential of using crowdsourced and social network data in socio-ecological investigations.

The methodological approach and data sources utilised in this chapter are now introduced. An analysis and subsequent discussion of results, in relation to the findings of previous research is then presented. Twitter, as a source of data for investigating reported seasonal variation in urban outdoor physical activity, is then evaluated before the chapter draws its overall conclusions.

4.3 Methodology

4.3.1 Study Area

Forty six urban green spaces were selected for study over a three month summer period (June - August 2015) and a three month winter period (December 2015 - February 2016). The urban green spaces were located in Birmingham, the second largest city in the United Kingdom with an estimated population of 1.1 million (ONS, 2014). Within the metropolitan area, there are nearly six hundred parks, public open spaces and nature reserves (BCC, 2016a), the most of any European city. Figure 2.4 (Chapter Two) depicts the locations of the urban green spaces in this study sample. The locations were chosen to reflect the diversity of urban green spaces found across the city and included parks and nature reserves of a range of sizes and with differing amounts of woodland, grassland, water and other characteristics. They offer a range of services and facilities and are found within different types of neighbourhoods.

4.3.2 Measurement of physical activity using Twitter

Created and launched in 2006, Twitter is a free microblogging service which enables users to communicate through short statuses and messages of up to 140 characters in length. Any registered Twitter user connected to the internet via a smart device or computer has the ability to receive and share information in real time. Twitter now reports 313 million monthly active users (Twitter, 2017) with over 500 million tweets uploaded per day (Internet Live Stats, 2017). This makes Twitter a highly influential player in the distribution of information and opinion (Mathioudakis and Koudas, 2010). Twitter data have already been used successfully in a diverse array of urban disciplines including land use classification (Frias-Martinez and Frias-Martinez, 2014; Zhan et al., 2014), environmental monitoring (Demirbas et al., 2010) and sentiment analysis (Hauthal and Burghardt, 2013; Klettner et al., 2013).

In this study, Twitter was used to create a corpus of tweets for investigation. English language tweets were downloaded via Twitter's REST API using the park names as queries. Duplicates and retweets were removed during pre-processing. Tweets were then manually screened and those referencing physical activity were collated into the dataset used herein.

This methodology created a corpus of tweets containing reference to physical activity being undertaken in urban green spaces in Birmingham. The tweets captured a range of information including the type of activity being engaged with, the weather conditions that exercise was undertaken in, who the Twitter user was exercising with, the length of time exercise was undertaken and mentions of notable events experienced during exercise. A small number of tweets from the corpus are presented in Table 4.1, demonstrating the variety of information the tweets contained.

Table 4.1 *A sample of the tweets in the dataset.*

Tweet	**Date of tweet creation**
"*Just cycled me and my legs up to Lickey Hills. Very Lickey up 'ere. #lickeyhills #birmingham #legs*"	08/08/2015
"*Head to Muntz Park for FREE #ParkLives Zumba (Fri 3.15pm)*"	23/07/2015
"Great football, great drama , great sportmanship , great day at the Aston Park Rangers Football Tournament today"	20/06/2015
"A muggy 8 miles through Sutton Park this morning, felt surprisingly good, after Tuesday nights tough race! #running #fitness"	25/06/2015
"Welcome to Billesley Common. It's blustery, it's cold, it's December rugby"	05/12/2015
"Lovely run in Summerfield Park with my son. What an awesome way to start 2016 #lovebrum #couchto5k"	01/01/2016
"Bird walk at Moseley Bog. We were listening to a black cap. #Urbannature #moseleybog #birdsong"	18/02/2016
"2 Superb Fitness Sessions this evening at Cofton Park with Coaches Matt & Dom"	11/01/2016

4.3.3 Measurement of meteorological variables

The procurement of atmospheric data at both high spatial and temporal resolutions over long periods of time remains challenging to urban climate researchers (Chapman et al., 2016). The discipline has begun to investigate the potential for citizen science and crowdsourcing for providing data appropriate for studies requiring atmospheric observations at high spatial and temporal resolutions (Muller et al., 2013; Overeem et al., 2013).

Netatmo weather stations (www.netatmo.com) provide a compromise between citizen science and crowdsourcing. The devices are commercially available to interested individuals who wish to monitor atmospheric parameters inside or outside a building. As a smart device, their in-built wifi connectivity enables the data each device collects to be transferred to a unified cloud server for storage where it is then made available for download through the Netatmo API. These devices are part of the ‘internet of things’ which is playing an increasingly important role in providing crowdsourced data (Muller et al., 2015). Netatmo weather stations were used to provide measurements of temperature (˚C) and rainfall (mm) across the study area. Measurements are taken by each device at a temporal resolution of 5 minutes and then uploaded to the Netatmo server infrastructure and made immediately available for download through the RESTful API. The closest outdoor Netatmo stations to the study sites were selected for study (Figure 2.4, Chapter Two). The position of these stations was sufficient to capture city-wide sufficient variability in the meteorological variables under study, despite low spatial coverage.

4.3.4 Statistical Analysis

After presentation of the relevant descriptive statistics for the datasets, the outputs of a number of statistical tests are presented to determine seasonal differences and any relationships present. Following a Shapiro-Wilk test for normality a paired sample t-test was used to determine seasonal differences in summer and winter weekday and weekend tweet frequencies. Descriptive statistics are presented for rainfall and temperature and the seasonal differences discussed. Spearman’s Rho are used to determine the relationships between a number of park attributes (park area, percentage tree cover, area of standing water, park amenities) and physical activity tweet frequency in order to better understand the spatial distribution and variability of physical activity tweets. Finally, Wilcoxon signed rank tests are used to determine the difference in tweet numbers between summer and winter and also the effect of organised sports events on this difference.

4.4 Results

A total of 2847 tweets were recorded in summer and 1920 were recorded in winter. Relative proportions show 59.8% of all tweets were received in summer and 40.2% in winter. Of these total tweets captured, 853 and 484 tweets were identified as relating to physical activity in summer and winter respectively. The relative proportions of physical activity tweets received in each season was generally reflective, but slightly more differentiated than the proportion of total tweets received in each season, with 63.8% received in summer and 36.2% in winter. The types of physical activities captured were highly diverse (Figure 4.1).

Figure 4.1 *A word cloud depicting the variety of physical activities reported in the received tweets. Word size is reflective of activity frequency, with larger activities appearing more frequently in the tweets.*

4.4.1 The influence of season on received physical activity tweets

Tweets mentioning physical activity occurred in more parks in summer than in winter. 29 parks showed a higher number of tweets mentioning physical activity in summer than in winter, 8 parks showed higher numbers of tweets in winter than in summer and 9 parks showed no change between seasons.

A Wilcoxon signed rank test identified a significant increase in tweets mentioning physical activity from winter to summer ($z=-3.418$m $p<0.005$). An effect size of -0.504 was calculated representing a medium change in the number of tweets mentioning physical activity between seasons according to Cohen's criteria for effect size.

With regard to the different types of activities recorded, the total number of activities occurring in the sampled locations increased from 19 in winter to 33 in summer, suggesting an increased variety in the physical activities taking place. As well as the increase in the number of physical activities taking place, it is also possible to identify a seasonal difference in the frequency that each activity is mentioned in tweets (Table 4.2). Using the number of tweets received in each season, six of the eight activity types showed an increase in tweet frequency from winter to summer. However, using normalised values, which identify differences between seasons with sample size variation taken into account, only three activities show increases from winter to summer; Active Parks activities (+239), team sports (+52) and fun sports (+25). Cycling and water sports remain proportionally the same in both seasons. A decrease in tweet frequency from winter to summer was observed for walking (-178), running (-61) and outdoor fitness (-27).

Table 4.2 *Seasonal differences in the frequency that each activity is mentioned in the received physical activity tweets. Normalised winter values are presented alongside actual tweet frequencies to account for differences in the sample sizes of total tweets received in winter and summer.*

	Winter	**Winter normalised values**	**Summer**
Active Parks activities	22	33	272
Running	150	222	161
Walking	223	331	153
Cycling	31	46	45
Water Sports	16	24	23
Team Sports	36	53	105
Outdoor Fitness	35	52	25
'Fun' Sports	1	1	26

Giving consideration to temporal variation in the number of physical tweets received over the course of a 7 day week, a paired sample t-test identified a significant difference between summer weekday and winter weekday received tweets ($t=-10.988$, $p<0.005$), however there was no significant difference found between winter weekend and summer weekend received tweets. The total number of tweets received per weekday in summer and winter followed a similar pattern (Figure 4.2). Decreases in tweet numbers occurred after the weekend which then increased to a midweek peak on Wednesday. Tweet numbers then fell on Thursday and Friday before increasing again at the weekend.

Investigations into the seasonal differences between when tweets were created showed early mornings and early evenings to be popular in summer compared to winter (Figure 4.3). The most popular time of creation for tweets in winter was found to be early and late afternoon.

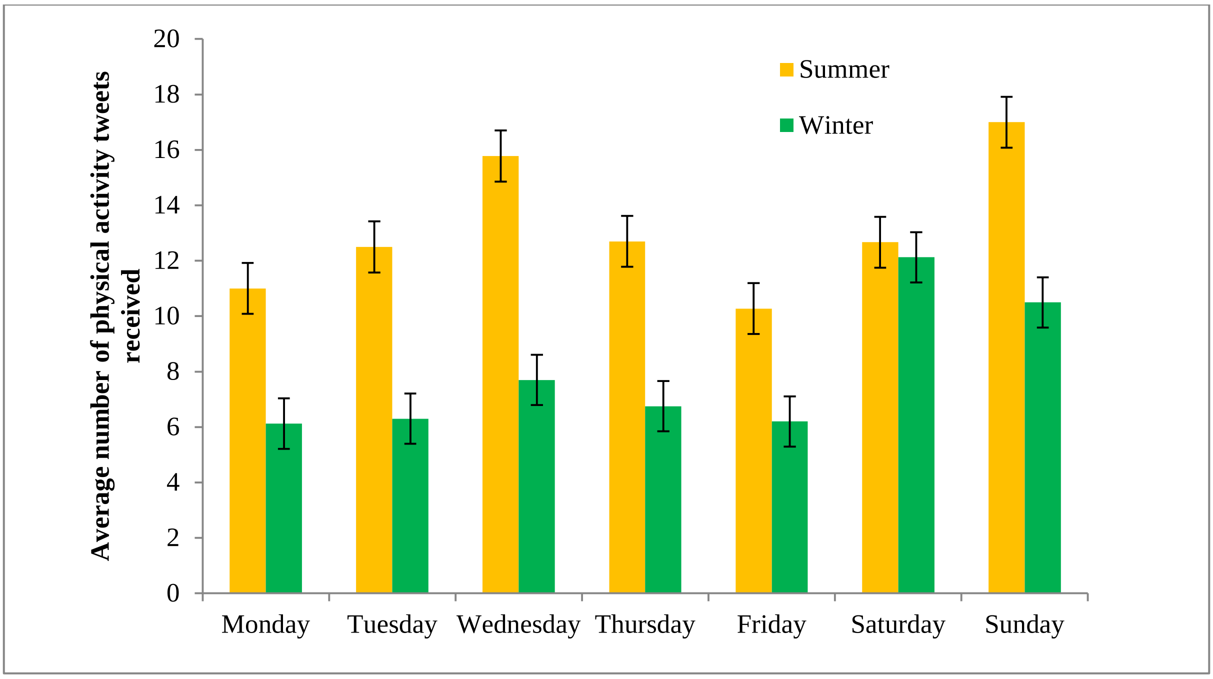


__Figure 4.2__ Seasonal differences in the average number of physical activity tweets received on each day of the week.

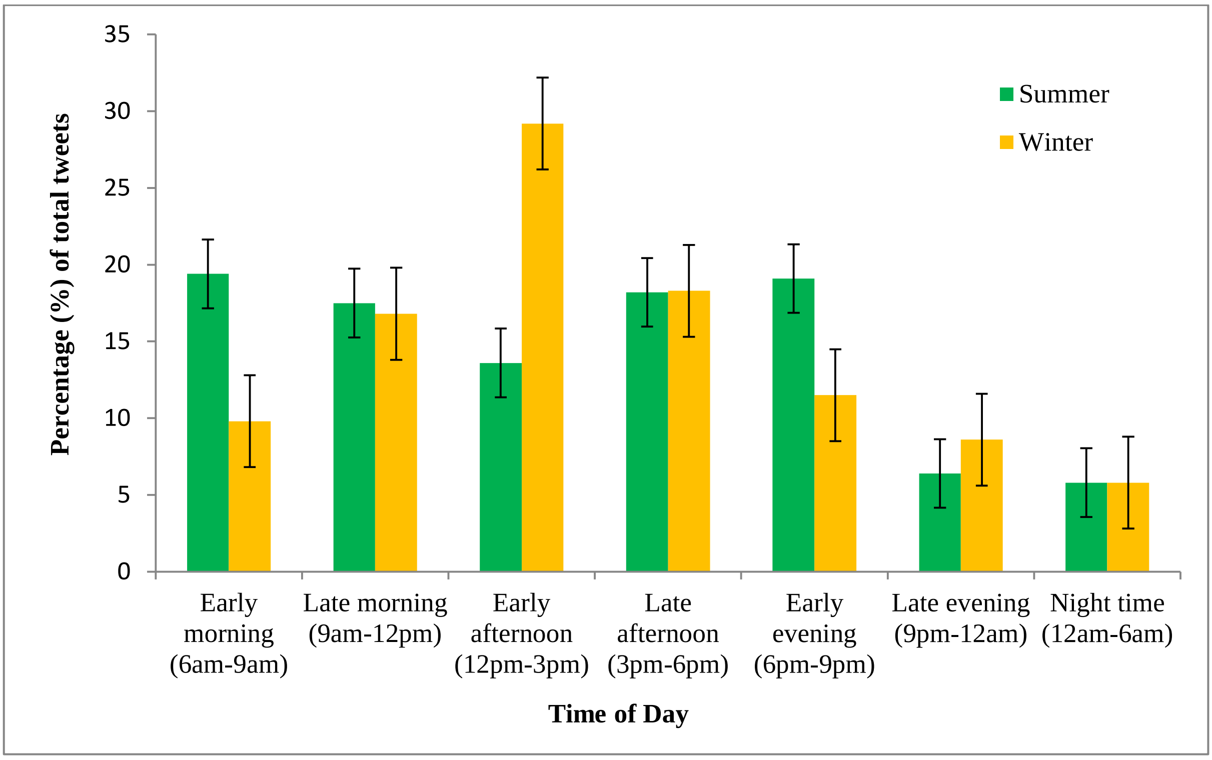


__Figure 4.3__ Seasonal differences in the number of tweets created at different times of the day.

4.4.2 The influence of meteorological variables on received physical activity tweets

To understand the role meteorology plays in the significant differences identified in the number of physical activity tweets received in winter and summer, each tweet was assigned a temperature value and rainfall binary (wet or dry). Using these hourly rainfall measurements, 92.9% of tweets were found to be created when no rainfall was recorded in summer, compared to 90.7% in winter. Both seasons showed significant numbers of tweets being created when no rainfall was recorded despite significant variation in mean temperatures at the time of tweet creation (Table 4.3).

Table 4.3 *Descriptive statistics for temperatures (°C) and presence of rainfall at the time of tweet creation for summer and winter.*

	Summer	**Winter**
Mean	17.6	6.6
Minimum value	4.8	-3.9
Maximum value	36.3	15.8
Range	31.5	19.7
Standard deviation	4.9	3.8
% of dry hours	95.6	95.6
% of wet hours	4.4	5.8

The difference between the temperatures at which tweets were created in summer and winter was less pronounced and a large range of temperatures were recorded for both seasons. When comparing the two seasons, summer temperatures were found to have a greater range than the winter temperatures at which tweets were created (Table 4.3).

4.4.3 The influence of park characteristics and organised sports events on received physical activity tweets

In order to understand better the spatial distribution and variability of physical activity tweets a number of park characteristics were considered including the park area, area of standing water, percentage tree cover and available facilities.

A spearman's rho showed that park area and the number of tweets mentioning physical activity, in both summer ($r_s[45]=0.181$, $p=0.230$) and winter ($r_s[45]=0.228$. $p=0.127$), were not significantly related. Similarly, there was no significant relationship between the area of standing water and the number of physical activity tweets for both summer ($r_s[45]=-0.059$, $p=0.699$) and winter ($r_s[45]=0.092$, $p=0.542$). The percentage of park tree cover was also insignificant for the number of winter tweets ($r_s[45]=0.188$, $p=0.210$), but significant for the summer ($r_s[45]=0.233$, $p<0.05$). Positive significant patterns were found between the number of park amenities relevant to physical activity and the number of physical activity tweets for both summer ($r_s[45]=0.480$, $p<0.05$) and winter ($r_s[45]=0.462$, $p<0.05$).

Whilst undertaking the categorisation of the received tweets it was clear that a number related to organised sports events (OSE). In order to ascertain if the presence of OSE in affecting engagement with physical activity, a comparison of tweets relating to independent physical activity and engagement with OSE was undertaken. Independent activity was classified as an individual or individuals engaging at their own time in any type of physical activity compared to OSE which were defined as pre-planned activities led by a specific individual or company. 70.6% of all physical activity tweets in summer were related to an OSE, compared to 45% of total physical activity tweets in winter.

Parks where only independent physical activity tweets were present occurred more in winter than in summer. 18 parks showed a higher number of tweets mentioning independent physical activity in winter than summer, 13 parks showed a higher number of tweets mentioning independent physical activity in summer than winter and 15 parks showed no change in the number of tweets mentioning independent physical activity (of which 14 out of the 15 detected no tweets mentioning independent physical activity). The opposite was true for OSE; with parks where only tweets about OSE occurred more in summer than in winter. 29 parks showed a higher number of tweets mentioning OSE in summer than in winter, 4 parks showed a higher number of tweets mentioning OSE in winter than in summer and 13 parks showed no change in the number of tweets mentioning OSE (of which 10 out of 13 detected no tweets mentioning OSE in both winter and summer).

A Wilcoxon signed rank test found the increase in tweets mentioning independent physical activity from summer to winter not to be significant. For OSE however, a significant difference was found in the number of tweets mentioning OSE in the summer compared to the winter. A significant increase in tweets was identified from winter to summer ($Z=-3.933$, $p<0.005$) and an effect size of -0.58 was calculated, representing a large change in the number of tweets mentioning OSE between the seasons based on Cohen's criteria for effect size.

Comparing the influence of meteorological variables on engagement with physical activity in winter, there was little difference in the mean temperature at which tweets relating to OSE or independent physical activity were recorded. A large difference was recorded however, between the percentages of tweets created when wet/dry (Table 4.4) with OSE demonstrating a higher proportion of tweets created in wet conditions.

Table 4.4 *Mean temperature (°C) and presence of rainfall at the time of tweet creation for independent physical activity and OSE in winter.*

	Organised Sports events	Independent physical activity
% created when wet	11.5	7.5
% created when dry	88.5	92.5
Mean temperature (˚C) at time of tweet creation	7.2	6.2

4.5 Discussion

Higher frequencies of physical activity tweets were received in summer compared to winter and an increased number of parks were shown to have people tweeting about engaging with physical activity in summer. While Twitter data are merely a proxy for physical activity engagement, and these observations are not representative of total engagement as not all individuals will tweet while taking part in physical activity; they do reflect the common notion that more people engage with outdoor physical activity in summer than in winter (Ma et al., 2006; Merchant et al., 2007; Tudor-Locke et al., 2004). This is explained by poorer weather conditions in winter and the effect this has on an individual's desire and motivation to engage with physical activity (Hug et al., 2009; Matthews et al., 2001). The continued, albeit lower, frequency of physical activity tweets in winter, is also consistent with findings from previous studies (Pivarnik et al., 2003; Uitenbroek, 1993).

When examining the seasonal differences identified between the types of activities, the higher diversity of activities reported in the summer is perhaps reflective of the increase of opportunities available for individuals to engage with. The decreases observed in the frequency of, for example, Active Parks activities from summer to winter is unsurprising. The observed lack of tennis, badminton, basketball and other court based activities reflects their conceptualisation as traditionally summer sports. A number of factors were considered in an attempt to find reasoning for the seasonal variation in physical activity tweets identified. The significant reduction of weekday tweets in winter compared to both winter weekend and summer weekday and weekend received tweets may be explained by variation in daylight hours (Cooper et al., 2010; Tucker and Gilliland, 2007). This notion is substantiated when consideration is given to the time of day that tweets were created. A large number of tweets were created in early mornings in summer reflecting changes in daylight hours capturing people using the early morning light to get out and exercise before work. In contrast, in winter, early and late afternoon were the most common time of day for physical activity. A plethora of studies have found that darkness inhibits outdoor physical activity in a variety of social groups, including low income women (Hoebeke, 2008), children of varying ethnicities (Brockman et al., 2011; Thompson et al., 2001) and the elderly (Bjornsdottir et al., 2012; Zimring et al., 2005). Self-reported reasons for this relate to a fear of crime and worries about personal safety (Bjornsdottir et al., 2012; Lee, 2005).

The influence of temperature on physical activity presented in previous studies is varied. Some studies identify temperature as a significant factor explaining the variance of engagement with physical activity between seasons (Nikolopoulou et al, 2001), while others describe a less obvious effect that is hard to differentiate from other influential factors (Humpel et al., 2002). In this study, temperature did not have a significant impact on physical activity tweet frequency for either season. There are a myriad of factors which can be influential in determining an individual’s response to cold weather, and the effect it has on their participation in outdoor physical activity, which makes generalising these relationships unwise. To highlight a few such factors, the role of clothing in mitigating low outdoor temperatures (Nikolopoulou et al., 2001; Thorsson et al., 2004) may help explain the lack of a significant impact of temperature on physical activity in both seasons, as well as the influence of habit in an individual's routine (Aarts et al., 1997) and their intrinsic motivation to remain active (Annesi, 2002) irrespective of temperature. The presence of park characteristics, such as shade, can also mitigate against high summer temperatures. While causality cannot be directly inferred, it may be that the positive association between percentage tree cover and physical activity tweets in summer is due to the presence of tree cover and thus shade, an important ecosystem service to individuals engaging with physical activity, cooling the environment and helping to mitigate against warm temperatures (Bastian et al., 2012). It may also be a reflection of the types of aesthetic environments that individuals like to engage with while they exercise and a tree filled space may be more appealing than spaces with different predominant landscape. Such observations could help urban planners create spaces with which individuals are more inclined to engage with and utilise. Indeed, there is potential for future research to make use of the multimedia attachments, which users commonly upload within their tweets, to examine the types of landscapes that are commonly engaged with for exercise.

The impact of rainfall was more pronounced with the majority of physical activity tweets in both seasons being created in dry conditions (Edwards et al., 2015; Tucker and Gilliland, 2007). While this is an expected observation, it is exactly because these observations are straightforward that Twitter provides a valuable source of data for empirical research into the effects of seasonality on outdoor physical activity. This finding extends to the different types of physical activities examined. With minimal exceptions, all showed similar proportions of tweets being created in wet or dry conditions, with the majority of tweets for all activities being created when dry. Caution is needed however, before making inferences from this data. It is not possible to know if individuals stop exercising in the rain or if they continue to exercise but refrain from tweeting about their activity because of concerns of getting their phones wet.

Concurrent with previous studies which find the presence of amenities to increase participation in outdoor exercise (Brodersen et al., 2005; Davidson and Lawson, 2006), the presence of certain amenities was found to influence the frequency of physical activity tweets. This is unsurprising given that some activities require certain facilities, especially team sports such as cricket, football, badminton, basketball, tennis, rugby and Frisbee, while cycling activities predominantly took place in locations with tarmacked paths or designated trails. Such consolidation of previous findings and the identification of expected outcomes again highlight the potential of Twitter data as a source of empirical data for urban green space research and in providing urban planners and with an evidence base for decision making. For example, Twitter data could be used to justify and monitor the implementation of a new cycle path. Twitter data have provided information on engagement with different types of physical activity and could be used to identify areas of low engagement with cycling where a path may help facilitate increased engagement. After implementation continued data capture could monitor cycling activity and provide evidence for any change in engagement with cycling.

It is also important to draw on the impact of organised sport events (OSE) on reported physical activity engagement and how this may explain the seasonal variation observed in some activity tweet frequencies. A variety of OSE took place in both the summer and the winter, contributing significantly to the role of urban green spaces as locations for engagement with physical activity. Twitter was found to be an important platform through which information about these OSE were shared. The continued provision of organised events through winter has previously been found to help increase engagement with physical activity (Sallis et al. 1998). Despite the observed reduction of OSE in winter in this study, the remaining OSE may explain the increased reporting of physical activity by Twitter users in less favourable weather conditions compared to independent physical activity. Indeed, for individual activities repeated precipitation events which are common in British winters, may decrease levels of physical activity for extended periods (Tucker and Gilliland, 2007), thus the provision of group activities may help motivate these individuals to take part regardless of the weather. Higher tweet frequencies of individuals participating in OSE when rainfall was recorded compared to tweet frequencies of independent physical activity, perhaps shows the importance of these events in encouraging individuals to participate in outdoor physical activity when they may typically be disinclined to do so. However, this cannot be directly inferred from the data as it is not possible to capture those engaging with independent physical activity or OSE and not reporting this in their tweets. Concerns about getting a smart phone wet may stop members of both activity groups tweeting in these conditions.

4.6 Evaluation of the use of Twitter data

This chapter has demonstrated that Twitter data can provide information on individuals' outdoor physical activity behaviours and used to investigate seasonal variation in physical activity in urban green space, adding to the results obtained using observational and recall-based methods reported in the literature. Twitter data were successfully used to investigate the reporting of engagement by individuals with physical activity in urban green spaces as well as providing insight into the range of activities with which individuals are participating. Twitter data provide greater spatial cover than some previous studies and the data collection is considerably less time and cost intensive than the methods traditionally employed in this field. The observations obtained from Twitter reflected well established notions of physical activity engagement and how this varies seasonally. This, and the significantly lower research cost of using Twitter data compared to traditional observational and subjective reporting methods makes it a valuable source of data for empirical research in this field. The chapter has demonstrated the utility of information provided in tweets in exploring the tweeting behaviours and outdoor physical activity practices associated with seasonal variations. Such information is helpful in providing a more nuanced understanding of the complexities of seasonal variation in engagement with outdoor physical activity in urban green spaces. It should be noted, however, that this work uses a relatively small dataset of tweets, thus caution is needed before inferring wide generalisations from these results. Larger tweet datasets should be collated to infer robust conclusions about physical activity variation in a particular green space or in a particular community. While this chapter demonstrates the method for obtaining such datasets, it is beyond the scope of this work and the dataset presented herein to make such inferences.

While this work has demonstrated the contextual information that Twitter can provide about the types of activities that individuals choose to engage, there is a limit to how informative 140 characters can be. There remain some significant limitations in the demographic information provided. For example, age (Uitenbrook, 1993) and ethnicity (Ma et al., 2006) have both been identified as influential factors in explaining the physical activity behaviours of individuals, information which is not available from Tweets. Without demographic information it remains challenging to establish the influential cultural dynamics operational in forming green space experiences.

It is also important to highlight the inherent biases in a dataset obtained from social networks given that the users of social networks do not reflect the diversity of the urban population they are investigating (Schwartz and Hochman, 2014). Not all members of the urban population will own a smart phone or actively engage with social networks. This has implications for investigating use of green space by specific sectors of the population. For example, older people (75+) show disproportionate levels of dis-engagement with these types of technologies (Zickuhr and Madden, 2012). A lack of specific demographic information may limit the utility of the information that Twitter data can provide in creating targeted policies aimed at increasing engagement with outdoor physical activity amongst certain age or ethnic groups.

While more investigation is needed to quantify the utility of tweet frequencies as a proxy for park visitation, the similar patterns found herein compared to previous studies may suggest (Ma et al., 2006; Merchant et al., 2007; Tudor-Locke et al., 2004) that Twitter data could be used to approximate park visitation and physical activity engagement, however, this would require a comparison between actual visit frequencies and received tweet frequencies to see if tweet data can accurately represent real visit data.

In order to promote year-round engagement with outdoor physical activity, it is necessary to understand the seasonal variation in such activity among the population (Ergler et al., 2016). The potential of providing information related to individuals' outdoor exercise behaviours to policy makers and urban planners is extensive (De Valck et al., 2016) and offers the possibility of enhancing the outdoor experiences of individual, and encouraging year-round engagement with outdoor physical activity in the green spaces they manage. This could be achieved through the development of interventions based around observations obtained from Twitter data. Planners could also capitalise on the promotive power of Twitter to encourage outdoor physical activity among its users, creating tweetable spaces, activities and challenges that users could engage with while exercising in urban green space.

Monitoring of council led physical activity programmes may also be possible through the capture of Twitter data and be a way to justify the continued running, or provide evidenced examples, of successful initiatives. For example, in Birmingham, the Active Parks and Parkslive programmes run by Birmingham City Council both provide opportunities for individuals to engage with a range of free physical activities in green spaces across the city. The #Parklives hashtag was often used by Twitter users detailing their engagement with these activities. By monitoring tweets containing this hashtag it would be possible for organisers to capture engagement, or use the other information users provide in their tweets such as who they are with or how they feel. Twitter could also provide a platform for organisers to actively engage with users to gain feedback on these sessions, by getting users to include specified hashtags in their tweets which could then be searched for through the Twitter API.

4.7 Conclusion

Seasonal variations in physical activity tweets between summer and winter were investigated. Significant seasonal variation in received physical activity tweets was identified and a number of factors were investigated in an attempt to explain this variation; including meteorological variables (rainfall and temperature), daylight hours, park characteristics and amenities and the presence of organised sports events.

The influence of rainfall was more significant than temperature on engagement with physical activity for both seasons. The presence of park amenities proved important in explaining the occurrence of a number of specific physical activities at certain locations, again in both seasons. Relationships between physical activity tweets and meteorology were found to be less significant than those presented in other studies. Other factors were found to be important in explaining the complex behaviour of engagement with physical activity, with the presence of organised sports events being particularly influential in increasing physical activity tweet frequencies in summer.

This chapter has successfully demonstrated the potential of Twitter data in investigating seasonal variation in physical activity in urban green spaces. Tweeting itself is a cultural practice and the information obtained from its users can inform investigations of seasonality in the social norms and cultural practices associated with outdoor physical activity. This method of data collection offers researcher a number of benefits over traditional observational and subjective reporting method, including a less time and labour intensive data collection process and greater spatial coverage. However, some limitations of this method still remain.

Little information about the demographic of individuals making up the dataset is available and until this is addressed it may limit the utility of results for policy makers.

While these limitations remain, this chapter has found Twitter data to be a useful addition to the methodological approaches employed to gather information on physical activity in urban green space. It offers a creative way to overcome methodological challenges associated with more traditional approaches and its successful employment consolidates the notion that geographic based approaches are well placed to investigate the relationship between seasonality and outdoor physical activity behaviours.

Chapter 5. The value of Twitter data for determining the emotional responses of people to urban green spaces: a case study and critical evaluation.[5]

Preface

Chapter Three of this thesis successfully demonstrated that the information in tweets can be used to inform investigations into how people use urban green spaces and the activities which take place within them. Chapter Four demonstrated that it is also possible to use Twitter data to investigate spatial and temporal variation in the activities with which people engage with in urban green spaces. However, urban green spaces are much more than purely functional locations. They are key components of the urban landscape which play a significant role in the creation of sense of place (Kim and Kaplan, 2004) for individuals and communities.

The importance of emotion in constructing everyday understandings of place should not be underestimated (Anderson and Smith, 2001). Thus, affordable and replicable ways of gathering emotional information on how individuals experience and feel urban green space is valuable information. An increasingly prevalent rhetoric in the planning literature cites that an understanding of 'sense of place' is important if planners are to successfully manage and develop urban landscapes to cater for the needs of the populations who live with within them (Jones and Evans, 2012). If this is to be taken seriously as an element of the social construction of urban landscapes which needs to be understood by planners for the most successful implementation of developments then, more emphasis needs to be placed on

[5] Roberts, H., Sadler, J. and Chapman, L. (2018) The value of Twitter data in determining the emotional responses of people to urban green spaces: a case study and critical evaluation, *Urban Studies*, 00 (0), 1-18. DOI: 10.1177/0042098017748544

understanding the existing place associations held by communities and individuals (Jones and Evans, 2012).

Emotional connection to place is intensely personal and visceral; making it hard to capture and highly changeable over both space and time (Jones and Evans, 2012). As such, it requires frequent monitoring which is not always possible using time and cost intensive methods such, as focus groups and neighbourhood surveys or questionnaires. This thesis posits that new data sources such as social networks are well placed to collect such emotional information. Indeed, being a free and publically available data source, tweets have the potential to provide cheap and accessible information on how individuals experience, perceive and feel as they navigate the urban landscape. Given that tweets are continually being created by users of the social network Twitter, they provide a continuous source of information which could be utilised by planner and academics alike to increase understanding of the emotional construction of urban environments.

This thesis takes the view that any positive emotional response or experience resulting from an interaction with an urban green space can be defined as a cultural ecosystem service. Research papers consistently underline the importance of the presence of green space for urban citizens but understanding the complexity in relationships between human-environment interactions makes it difficult to study the potential benefits such interactions may foster (Kabisch et al., 2015). As a result, studies tend to point to the ‘potential’ benefits of such interactions and new methods are required to enable the quantification of these interactions and the emotional responses associated with them. Chapter Five presents a first attempt to capture and quantify the emotional responses of individuals as they report their use of urban green spaces on Twitter. In doing so it addresses the following objective: to evaluate the use of such data in investigating how individuals respond to and experience urban green space.

5.1Abstract

Interactions between humans and nature are understood to be beneficial for human well-being. In cities, urban green spaces provide many benefits to urban populations in terms of mental and emotional well-being. Through a case study of sixty urban green spaces in Birmingham, United Kingdom, this chapter investigates spatial and temporal variation in the emotions experienced by individuals whilst they use and interact with urban green space. Using crowdsourced information as the basis of emotional explorations, sentiment analysis is performed on over 10,000 tweets to ascertain the positivity/negativity of individual emotions. Positive responses were found to be more common than negative responses across all seasons, with happiness and appreciation of nature being the most common positive emotions identified. For the negative responses, fear and anger were present in similar amounts with smaller annotations of sadness and disgust also present. Negative responses were found to be variable across seasons and closely related to the events and occurrences happening in the urban green spaces. The findings presented in this chapter concur with existing research, suggesting that crowdsourced information from Twitter offers a new way to successfully progress methodological approaches in understanding human interactions with urban green spaces. The information these data can provide has utility for urban planners and park management providing insight into the features of urban green spaces that enhance positive outdoor experience.

5.2 Introduction

5.2.1 Green space and well-being

Contact with the natural environment is understood to be a fundamental component of human and societal well-being (Wilson, 1984; Chiesura, 2004; Miller, 2005; Fuller and Gaston, 2009). The potential benefits offered to human populations from natural environments are increasingly significant in an urbanised society where green spaces are under threat by densification of urban form (Dallimer et al., 2011). The benefits that nature and green space can provide to human populations are a vital component of ecosystem services (Ehrlich and Ehrlich, 1981; Costanza et al., 1997; Daily, 1997; MEA, 2005; Elmqvist, 2011).

Green spaces in cities provide the opportunity for people to connect with nature (Kremer et al., 2016), a need thought to be innate to all human individuals (Wilson 1984; Kellert and Wilson 1995). Two current schools of thought co-exist in attempting to explain why contact with nature is beneficial; the first hypothesises that such experiences have a direct effect on the nervous system, resulting in stress reduction and attention restoration through reduced exposure to environmental stressors (Kaplan, 1995; Honold et al., 2016). In urban areas, stressors such as noise and crowding are key motivators encouraging individuals to seek out contact with more natural environments (Hartig and Staats, 2006; van den Berg et al, 2010). The second, based in psycho-evolutionary theory, suggests that positive responses to nature may have a genetic basis given the adaptive significance this would have had throughout human evolutionary history (Ulrich, 1993; Bratman et al., 2012). It has been theorised that human appreciation and enjoyment of the natural world, termed *biophilia*, has developed as a consequence of our evolutionary trajectory (Kellert and Wilson, 1995).

Urban green spaces also provide space for social interaction within the cityscape. From a societal processes point of view, green spaces can facilitate social cohesion and be a place in which social ties are created and maintained (Kweon et al., 1998; Maas et al., 2009); factors which have been associated with improved well-being among urban citizens (Lee and Maheswaran, 2011).

The concept of subjective well-being (SWB) provides a link between well-being and emotion, encompassing an individual's emotional responses, domain satisfactions and life satisfactions (Diener et al., 1999). Emotion is a central part of being human; influencing an individual's notion of self (Davidson and Milligan, 2004). Understanding the cues leading to particular emotions being experienced is thus a significant area of research. However, the study of human emotion has long been challenging to those engaging with it as the exact definition of 'emotion' is still debated (Gedron, 2010; Scheer, 2012). Biological conceptualisations of emotions suggest them to be a result of bodily arousal and cognitive processes which lead to the release of chemicals in the brain (Davidson and Sutton, 1995; Lane and Nadel, 2002). These chemical signals alter the neural circuitry used leading to a change in experienced emotion in the individual (Pert, 1997) thus completing the process of chemical changes causing affective change. Six basic emotions have been defined; happiness, sadness, anger, fear, surprise and disgust (Ekman and Friesen, 1971; Ekman, 1992), and differentiated from other affective phenomenon which are more socially or culturally influenced (Ekman, 1999).

Despite the uncertainties and complexities inherent in studies of SWB and emotion, it is increasingly accepted that an individual's surroundings are influential on the emotions they experience through the stimuli they present (Oschsner and Gross, 2007). Such a view can be seen as an attempt to combine biological and cultural approaches to studying emotion; in that an individual's environment is viewed as a trigger for certain responses and emotion. Indeed, it is widely accepted that the process of generating emotion begins outside of conscious awareness (LeDoux, 1996). The notion that the physical locality of a space and the cultural context in which the interaction occurs, are influential on resultant emotions (Ulrich, 1983; Davidson and Milligan, 2004) provides a useful viewpoint in exploring how we make sense of a space.

Emotional responses are culturally influenced (Lutz and White, 1986; Kövecses, 2003). Indeed, the way emotions become spatially located is highly dependent on localised social and cultural constructions of place (Kahn, 1999). This acts to consolidate how a place is imagined and has implications for the embodied experiences of those who visit. The relationship between people and nature is a particularly culturally sensitive narrative (Simmons, 2013) with landscapes often reflecting the self-definitions of the individuals within a particular cultural context (Greider and Garkovich, 1994). The urban landscape is being increasingly experienced by individuals as global urban populations continue to expand (UN HABITAT, 2016) making it an important landscape to study in terms of the effects it may have on how individuals feel within it. Urban living offers comparatively fewer opportunities for interaction with nature and natural environments (Soga et al., 2015), yet the demand to interact with, and experience, nature remains prominent within urban communities (Ellard, 2015; Soga et al., 2016). Given the positive implications of human interactions with nature and green spaces it is logical to assume that this reduction in opportunity for interaction may have negative implications for individual well-being. Indeed, studies have pointed to a disconnection with nature leading to unhealthier and unhappier individuals (Feral, 1998; Miller, 2005; Soga and Gaston, 2016).

5.2.2 Approaches to investigating associations between green space and well-being

Previous research investigating the effects of natural environments on well-being has taken two lines of enquiry, relying on observational and experimental evidence. Observational studies have sought generalised associations between well-being, in its broadest sense, and a number of environmental characteristics, including proximity of homes to natural environments (Wells and Evans, 2003; van den Berg et al., 2010) and local environmental quality (Brereton et al., 2008).

Experimental studies have dissected the components of well-being, considering emotional response, levels of happiness and mood in their investigations. A number of these studies have indicated that contact with nature promotes positive emotional responses (Mayer et al., 2008; Nisbet and Zelenski, 2011; Tsunetsugu et al., 2013). A recent meta-analysis of empirical research into the effects of natural environments of emotional well-being undertaken by McMahan and Estes (2015) provides a number of useful findings. Across the studies included, contact with natural environments was consistently associated with higher levels of positivity and lower levels of negativity, suggesting even brief contact may have substantial benefits on well-being. Observational and experimental approaches, whilst providing valuable insight into the relationship between green space exposure and well-being, have several limitations associated with them. Findings from experimental studies are difficult to extrapolate to real-world contexts, whereas observational approaches are time and cost intensive and as a result often lack replicability or focus on short periods of time.

As the use of digital technologies increases, there is recognition among geographic disciplines that they are transforming the production and experience of space, place, nature and landscapes; and how geographical research is conducted (Ash et al., 2016). Within urban research the use of technological infrastructures, particularly smart phone applications and social network data, are increasing the spatial and temporal coverage of traditional monitoring networks (Resch, 2013) and is emerging in a variety of contexts; including air quality monitoring (Postolache et al., 2009), sound-scaping the urban landscape (Aiello et al., 2016), studying patterns of human behaviour (Sagl et al., 2012) and improving the efficiency of urban infrastructures (Hancke et al., 2013). The rapid emergence of social networks offers great potential to enable the extraction of human-centred information about a location through technological means, contributing to the relatively new concept of augmented space (Cresswell, 2004).

The development of technologies, such as smart devices, that support social networks provides researchers with a wealth of information about individuals. Indeed, data derived from social networks have already been utilised in a number of applications to support a range of urban inquiries. For example, Instagram, a photo-sharing social network, has been employed to operationalise key concepts in the study of the city, such as sociospatial patterning (Boy and Uitermark, 2016), and inequalities (Shelton et al., 2015); while geolocation based Foursquare has been employed to investigate socialisation in space (Sutko and Silva, 2011) and dynamic flows of people between discrete urban locations (Silva et al., 2014). Twitter, the social media of relevance to this chapter has also been successfully employed to investigate urban landscapes, including land use categorisation (Frias-Martinez and Frias-Martinez, 2014) and the spatial variability of crime (Yardi and Boyd, 2010).

Created and launched in 2006, Twitter is a free microblogging service enabling users to communicate through short statuses of up to 140 characters in length. Twitter users can receive and share information making Twitter a highly influential player in the distribution of information, opinion and emotion (Mathioudakis and Koudas, 2010). The emotional responses of people to space vary depending on how they experience and the environment around them; particularly in terms of their sensory perception and the subsequent feelings this produces (Rodaway, 1994). Such a notion is central to the investigation of emotion in urban areas. Crowdsourcing of information from Twitter is now being considered for its potential in exploring the emotions experienced and reported in urban spaces, as demonstrated by Klettner et al. (2013) who found the environmental characteristics of an individual's surroundings to be influential in their affective response. Spatio-temporal variations in crowd mood have also been studied through the categorisation of semantic terms from tweet text (Wakamiya et al. 2015). Twitter thus provides complementary information that can be used alongside biosensing and subjective reporting (Resch et al., 2015).

The studies presented above highlight the potential of Twitter data in providing sentiment information in generalised urban areas, but to date, little research has made use of Twitter data to investigate emotions related to urban land use and none exists for urban green space. Understanding how people feel in the green spaces they experience, and identifying the features which elicit specific responses, can provide urban planners and managers with an evidence base from which spaces can be created which enhance the experience of users. The notion that the physical environment, specifically green spaces, has a significant effect on human emotion while an individual experiences them falls within the remit of a number of geographical and psychological disciplines whilst not being fully embraced by any. The experience of place and its effect on the individual is central to theories of place (Shamai and Ilatov, 2005) with the effects of environmental characteristics and human use of space playing an important role in understanding the development of place attachments (Stedman, 2003; Patterson and Williams, 2005). This chapter synthesises these parallel themes to create an inter-disciplinary piece of research examining the effect of urban green spaces on experienced human emotion. The objectives are threefold:

(1) To introduce Twitter as a source of data for urban emotion research and explore its potential in capturing the emotions experienced by individuals while they are in urban green space;
(2) To present a case study in which the spatial and temporal variation of emotional responses to urban green space is investigated; and to explore the potential of Twitter data in explaining why these emotional responses are observed;
(3) To evaluate Twitter as a source of data in this context.

The focus of this study is on the emotional responses of people to urban green space and does not attempt a comparison of these results with emotional responses to other urban spaces (shopping centres, restaurants etc).

5.3 Methodology

A case study is now presented in which the application of Twitter data to understand the emotional responses of individuals to urban green spaces is explored.

5.3.1 Study Area

The urban green space study sites are located in Birmingham, the second largest city in the United Kingdom with a population of approximately 1.1 million people (ONS, 2014). Within the metropolitan area, there are nearly 600 public parks, open spaces and nature reserves, the most of any European city (BCC, 2016a). They provide an important resource for the surrounding populations in terms of their contribution to cultural ecosystem service provision. A total of sixty urban green spaces in Birmingham were included in this study (Figure 2.1, Chapter 2).

The parks used in this study were chosen to reflect the diversity of urban green spaces found across the city. A cross section of spaces were chosen including parks a range of sizes and with differing amounts of woodland, grassland, water and other green attributes. They offer a range of services and facilities and are found within different types of neighbourhoods. Alongside forty six parks, a number of green linear features were also included for investigation consisting of the footpaths along four rivers and seven canals, and three cycle ways.

5.3.2 Tweet corpus creation

Twitter provides publically accessible data through its REST API. R, and specifically the 'twitteR' package were used as the interface through which connection to the Twitter API was made. The 'twitteR' package is designed specifically for working with the Twitter API and has the necessary coding functions already in place. Connecting to the Twitter API using this procedure makes use of the OAuth protocol, a method which enables third party researchers to access user data without gaining access to passwords and other private information (Hawker, 2010; Russell, 2013), thus ensuring Twitter user confidentiality. Access through OAuth grants a third party user an access token and an access token secret which act as their credentials to access the user data.

English language tweets were downloaded approximately every ten days from the API to ensure maximal temporal coverage over a period of twelve months, from June 2015 to May 2016. This frequency of tweet download was used as the REST API keeps tweets available for download for approximately 7-9 days. More frequent downloading would thus result in the unnecessary capture of duplicate tweets. A search query was used to ensure that the tweets downloaded were related to one of the sixty sites included in the study. The search query used was the name (or names) of the green space, for example 'Cannon Hill Park'. This provided a simple, effective and replicable way which ensured only tweets containing references to the specific green spaces included in this sample were captured.

5.3.3 Annotation

Once downloaded, tweets were assigned by one manual annotator into one of three categories: positive, negative or neutral. This annotation was based on the presence of emotive words, emoticons or meaning in the tweet text. Where images were uploaded alongside a tweet, these were used to aid annotation and provide the annotator with more context. Subsequently, the positive and negative annotated tweets were further categorised into distinct emotions. The high level emotions chosen included Ekman's six basic emotions (anger, disgust, fear, sadness, happiness, surprise) plus love and calmness, in line with previous research which has undertaken sentiment analysis of Twitter data (Roberts et al., 2012; Resch et al., 2015). These emotions are arranged into the ontology shown in Figure 5.1. Beauty was included as a sub-category to the positive tweets but outside of the emotions to account for the large amount of tweets referencing the beauty of nature and the landscape. Each tweet could only be assigned into one of these emotion categories based on the strongest present emotion.

One researcher was used to annotate the majority of tweets in the sample given the time required for the task. However, four other annotators were used to annotate a random sample of 1,000 tweets, in order to assess reliability between different human annotators. A metric of comparison was derived (K=0.666) suggesting sufficient inter-annotator reliability (Landis and Koch, 1977).

Each tweet was also categorised by user type: male, female or organisation. Male and female categories reflect the gender of the Twitter user, while organisations include the private companies, local council initiatives and local businesses found to be captured in the sample.

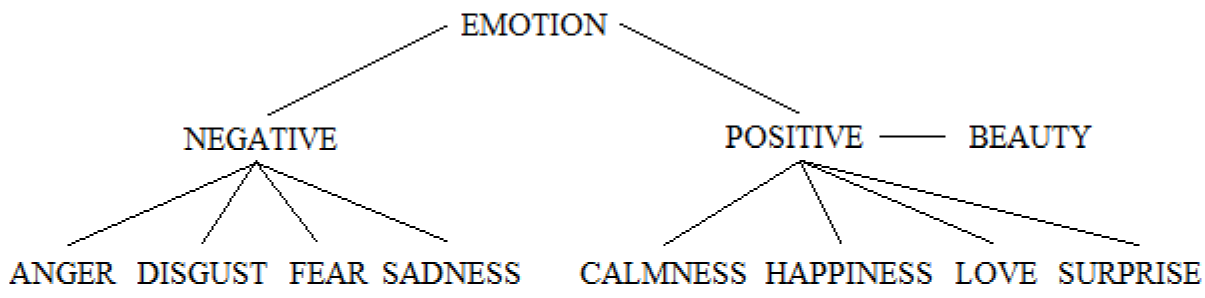


Figure 5.1 *High level emotion ontology for the emotions used in tweet annotation.*

5.3.4 Analysis

After discussion of the relevant descriptive statistics for the dataset, the outputs of various analyses to examine spatial and temporal variation in the positivity of tweets are presented. In order to assess spatial variation in tweet positivity over the twelve month period, frequencies of positive and negative tweets for each green space in the sample study are mapped. Temporal variation was investigated by first presenting descriptive statistics for each season (Summer: June July August, Autumn: September October November, Winter: December January February, Spring: March April May). The difference in tweet numbers between the seasons was analysed using Friedman tests as data violated the assumption of normality.

Descriptive statistics are presented for the high positive and negative emotions, before a thematic analysis to identify recurring factors influencing the expressed responses of Twitter users in the dataset.

5.4 Results and Discussion

5.4.1 Positivity, Negativity and Neutrality of tweet responses

A total of 10,197 tweets were downloaded relating to the sixty study sites over a twelve month period. Of these tweets, 68.4% could not be assigned either a positive or negative association and make up the neutral category. These neutral tweets consisted of news reports, advertised events or individuals stating to be in a location but with no identifiable emotion present in the tweet. It is not uncommon for the majority of tweets in a dataset used for sentiment analysis to be defined as neutral, based on a lack of positive or negative leaning, and removed from subsequent analysis (Roberts et al., 2012). The remaining tweets show higher levels of positivity than negativity, 24.6% and 7% respectively; concordant with the findings of studies which have relied on experimental approaches, in which higher levels of positive sentiment are consistently observed from people in nature (Mayer et al., 2008; Nisbet and Zelenski, 2011; McMahan and Estes, 2015).

These findings corroborate with the notion that the beneficial effects of nature on well-being are driven by increases in positivity. Indeed, from an evolutionary perspective, increases in positive emotions experienced while in natural settings was likely to be adaptive through much of evolutionary history given the presence of beneficial resources in natural environments. Thus the primary function of positive emotions would be to encourage engagement with this evolutionary adaptive behaviour (McMahan and Estes, 2015).

Organisations such as private companies, local council initiatives and local businesses were found to make up the majority of positive and negative tweet responses (44.6% (Figure 5.2). Total responses from male users were found to be slightly higher (29.9%) than total responses from female users (25.5%). Male and female users had similar numbers of positive responses, while male users showed a slightly higher number of negative responses.

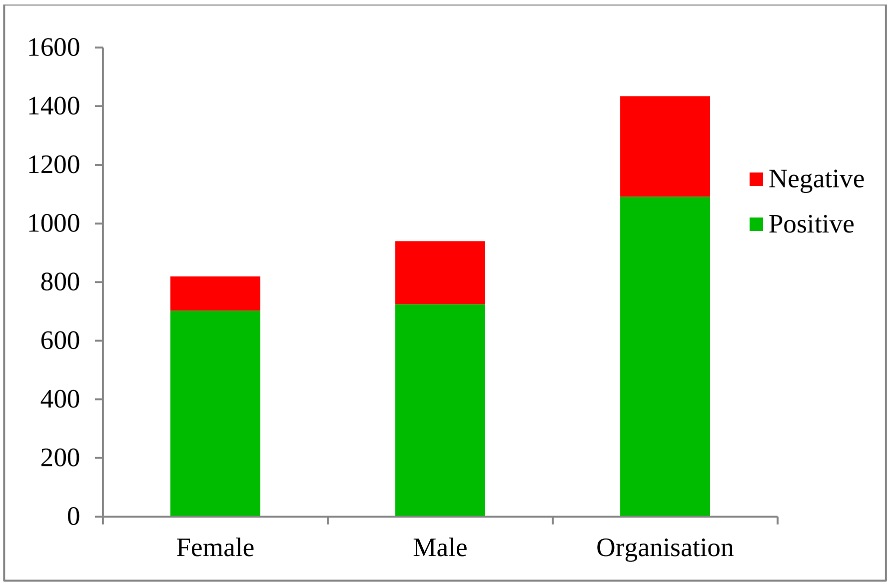


Figure 5.2 *A comparison of the number of positive and negative tweet responses for organisations and male and female users over the twelve month study period for all sixty study locations.*

In terms of spatial variation, the proportions of positive and negative tweets received from the discrete urban parks were highly variable (Figure 5.3), again corroborating the findings of previous study. Indeed, Bertrand et al. (2013) found sentiment to vary significantly at fine grained spatial scales within a cityscape.

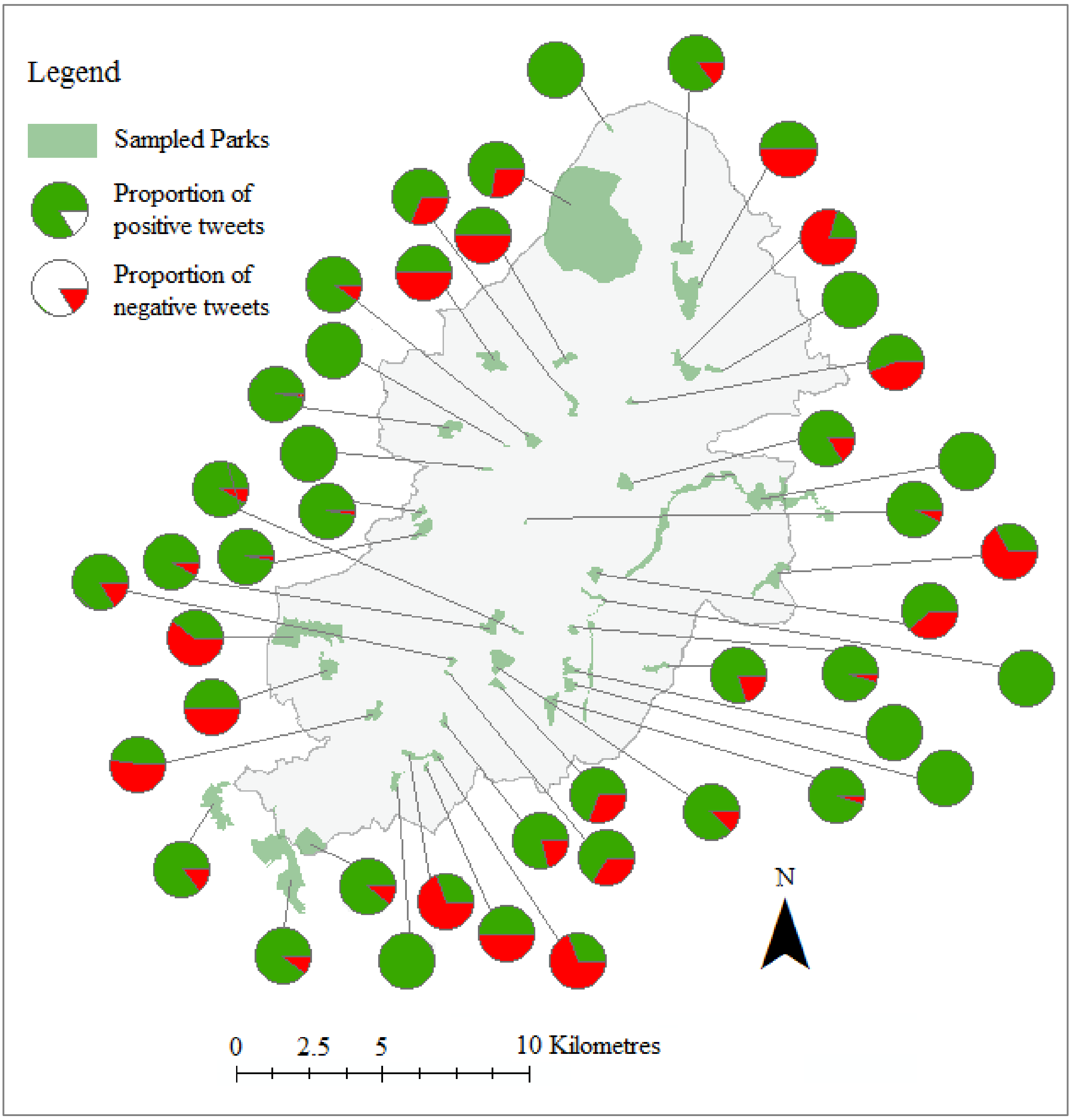


Figure 5.3 *The proportion of positive and negative tweets received per urban park over the twelve month study period (cycle paths, rivers and canals were excluded). Numbers detail the total of positive and negative tweets received from park.*

Examining each space makes it is possible to identify the causes of these positive and negative responses. For example, the emotional responses captured in Moseley Bog and Hillhook Reserve were all positive; largely the result of individuals enjoying the natural surroundings of the nature reserves, particularly with family. In contrast, the majority of responses in Pype Hayes Park and Woodgate Valley Country Park were negative. In the first case, this was found to be caused by the prolonged presence of the travelling community on a number of occasions throughout the year, as well as by anger at the decision by the council to no longer hold a large fireworks display at this site in November. In the latter case, negativity was due to fly tipping and the presence of litter and used drugs paraphernalia. In both these positive and negative examples, this chapter demonstrates that it is possible to identify the cause of an emotional response using Twitter data, which presents a useful resource for urban planners and park management. Identifying the cause of negative responses such as litter, can be used to ensure effective utilisation of limited council resources and targeted improvement initiatives. Identifying locations where people enjoy and interact with nature may help to ensure and justify their continued presence in the urban landscape, when pressures for development can be intense. The ability to identify the causes of emotional responses, afforded by Twitter, would enable evidence based management in these and other locations, across the city.

Temporal variation was also identified over the twelve month period, in the number of total tweets received within each category (Figure 5.4). Friedman tests show a significant difference in the total number of positive tweets received in each season ($\chi^2(2)=10.826$, $p=0.013$) and the total number of negative tweets received in each season, although this relationship was less strong ($\chi^2(2)=7.979$, $p=0.046$). Winter differs more from the other seasons (Figure 5.5) and is characterised by significantly higher levels of negative tweets and lower levels of positive tweets. This is perhaps surprising given that previous research has found natural settings in the urban landscape to remain favoured places by urban inhabitants in winter (Korpela, 2003). This suggests a need for the development of a replicable method to engage with and investigate urban green spaces at this scale, such as the one proposed in this chapter.

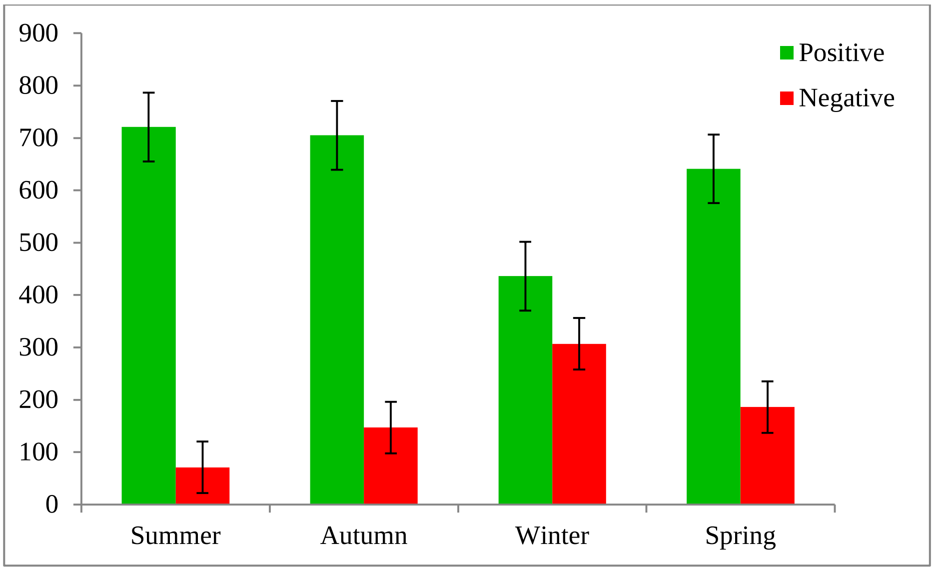


Figure 5.4 *The total number of tweets assigned to positive and negative categories for each season in the twelve month study period.*

5.4.2 Emotional response and thematic analysis

Over the twelve month period the most common positive emotions identified were happiness and beauty followed by love, calmness and surprise (Figure 5.5).

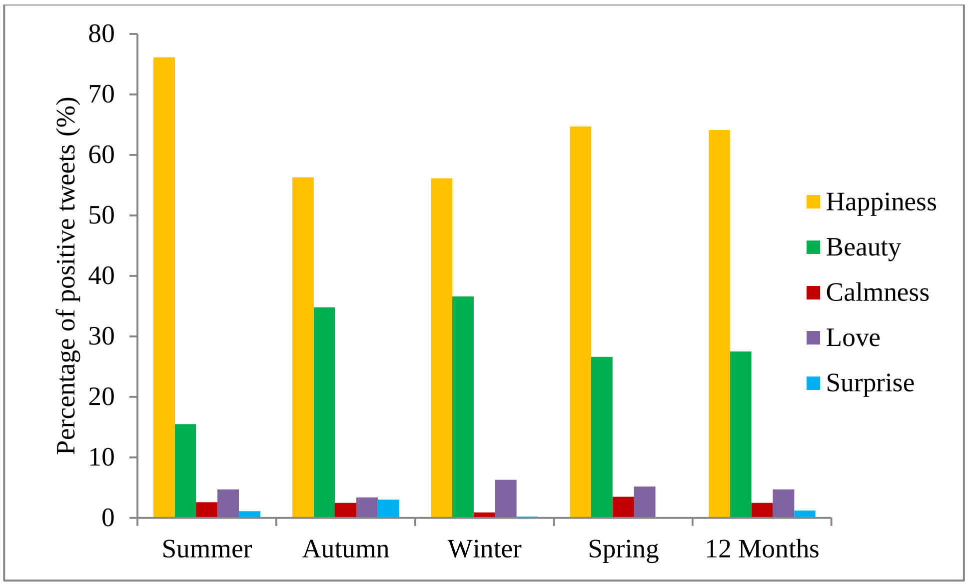


Figure 5.5 *The percentage of positive tweets assigned into the five positive emotions over the twelve month period and in each season.*

Happiness was present in the highest frequencies in each season, consistent with previous studies reporting happiness as a key response to green space and nature (MacKerron and Mourato, 2013). On examination of the tweets, happiness was assigned to a variety of tweets ranging from reports of general good mood, having a good day, enjoying or looking forward to an event or activity and having fun with friends. Happiness responses were thus elicited as responses to events and activities occurring in the parks as well as people reporting being happy in parks. For this reason, emotional information collected through Twitter data could be used as a monitoring tool for interventions or events. For example, in this study many happiness responses were the result of an individual engaging with an Active Parks activity. These events are provided free through council and commercial sponsorship and the initiative requires evaluation to establish whether it should be continued. The positive responses to these events could be used to justify continued funding. While not an emotion as such, beauty was added to capture positive responses of people to the spaces themselves. Reponses in this category were often found to be associated with physical activity, with people tweeting about the views and natural things they encounter as they exercise outdoors. Beauty responses were most common in autumn and winter (Figure 5.6), explained by the enjoyment of individuals in the autumnal colours and scenery, while frost was popular in winter. Responses to beauty in urban green spaces and nature speak biophilia hypothesis; the notion that people enjoy and respond positively to natural surroundings (Wilson, 1984; Ulrich, 1993; Kellert and Wilson, 1995). They also capture environmentally induced awe and appreciation on the landscape, which as a phenomenon, has been discussed in previous research (Davidson et al., 2005).

The final three positive emotions; love, calmness and surprise were all present but to lesser extents. The presence of love can be explained in two ways, being present in tweets mentioning the love of green space, or the love for someone they are in the green space with. The detection of calmness is consistent with findings from a previous study in which tranquillity was found to be a characteristic associated with more natural settings (Herzog and Chernick, 2000). Finally, surprise was present in association with notions of a green space or activity being better than expected. This is reflective of previous findings which identified repeated cases of surprise at the unexpected positive experiences urban green spaces were found to provide (Rishbeth and Finney, 2006).

For the negative emotions, fear and anger showed similar frequencies over the twelve month period, despite some seasonal variation, followed by sadness and disgust (Figure 5.6).

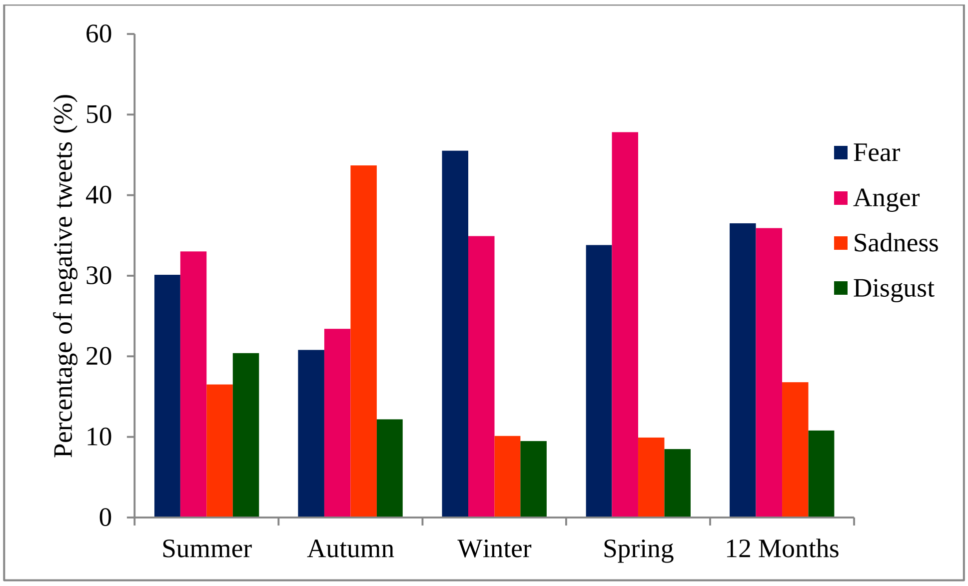


Figure 5.6 *The percentage of negative tweets assigned into the four negative emotions over the twelve month period and in each season.*

In comparison to the tweets annotated with positive emotions, the negatively annotated tweets were largely associated with events occurring within the green space, rather than a response to the green spaces themselves. For example, the sadness peak observed in autumn was caused by the responses of individuals to the accidental death of a woman whilst she visited one of the sites under study. The remaining sadness responses across all seasons could be attributed to losing at sports events, the cancellation of an event or poor weather.

In contrast to a number of previous studies which have suggested urban green space as a remediation to stress and related emotions such as anger (Groenewegen et al., 2006; Tyrvainen et al., 2013) this study identified a number of angry tweet responses to urban green spaces. The occurrences eliciting anger response were largely criminal or anti-social behaviour such as incidents of personal theft, vandalism, aggressive dogs and excessive noise. The high level of anger responses in spring was attributed to council plans to remove play areas from a number of sites in the study sample. Disgust across all seasons was found to be a response to varying acts of anti-social behaviour including graffiti, fly tipping, dog fouling and litter. The exception of a negative emotion being a response to occurrences within the space rather than the space itself was fear. Fear was commonly presented as a response to not feeling safe in a location, particularly at night. Such responses were present throughout the year.

The identification of negative responses from Twitter data show the generalised opinion of green space providing positive experience should be challenged and that emotional responses to space are highly spatially and temporally variable. Indeed, both fear and disgust have been identified as emotions experienced by urban individuals when they experience green space and nature (Bixler et al., 1994) and the findings of this chapter continue to support the notion of these emotions being articulated by some individuals when they come into contact with green space. Understanding the negative responses of individuals to natural environments and their causes is important if the barriers to use and enjoyment of urban green spaces are to be fully understood and addressed (Bixler and Floyd, 1997) and spaces created to enhance positive outdoor experiences and improve well-being.

From this analysis one can identify a number of recurring themes leading to the resultant emotions being expressed by individuals in their tweets. For the positive tweets, distinct themes are evident which resulted in positive emotions being expressed (Table 5.1). For the negative emotions however, general themes were harder to identify. In contrast, the negative emotions reported in tweets were often the result of specific events to which individuals were reacting to. For example, these events included criminal behaviour (muggings, sexual assaults, arson and evidence of drug use), conflict between users (cyclists and walkers), poor weather limiting planned physical activity, anti-social behaviour, fungal infection of local animal populations and the presence of the travelling community.

Table 5.1. *Recurring themes identified causing a positive response.*

Theme/cause of positivity	**Associated emotion**
Socialising and enjoying time with friends	Love/Happiness
Enjoyment of nature and the landscape	Beauty/Happiness/Calmness
Enjoyment of a music event	Happiness/Love/Surprise
Enjoyment of a sports event or exercise	Happiness/Love/Calmness

5.5 Evaluation of using Twitter data to gather information about emotional responses to urban green spaces

This chapter has demonstrated the how Twitter data can be used to gain insight into emotional responses of individuals to urban green spaces. The findings indicate that tweet text can be used for sentiment analysis and subsequent corpus analysis to ascertain the positive and negative responses of individuals to urban green spaces, as well as identify a number of distinct associated emotions.

Aesthetic appreciation of the natural landscape is a key cultural ecosystem service defined by the MEA (2005) and has been captured successfully in this chapter through the identification of tweets responding to the beauty. Thus Twitter data have potential for investigating this ecosystem service and how individuals respond to the natural landscapes they experience.

The use of Twitter data has a number of other benefits when compared to previous methods employed to this end. For example, the constant creation of data by Twitter users and the freely accessible nature of Twitter data through the REST API provides opportunities for studies with greater spatial and temporal coverage than has previously been undertaken. Data collection through Twitter is unobtrusive and less time and cost intensive in comparison to observational and experimental methods and can be used to develop replicable and standardised analytical procedures.

The identification of findings concurrent with previous research into the emotional responses of individuals to urban green space suggest that this method is a viable way to progress methodologically. This information is valuable to urban planners and managers and has the potential to identify features which enhance the outdoor experience of urban green space users.

However, the use of Twitter data has a number of limitations that should be addressed if this method is to be employed most effectively in future research (Roberts, 2017). There remain limitations to how informative 140 characters can be; especially in relation to emotions, which are highly complex, subjective experiences. In tweets where no explicit reason is provided and given that no wider context to each tweet is available, it is not always possible to ascertain why an emotion is being experienced. This is particularly relevant to studies of emotional responses to space. For example, in this chapter issues of causality were identified in the detection of love from the tweet text: it was sometimes unclear whether this response was due to the green space or to companions or an activity being engaged with in the space.

Additionally, the inherent biases in a data derived from social networks should be considered (Hannay and Baatard, 2011). No up-to-date figures exist for the number of UK Twitter users, making the ratio of users to non-users an unknown quantity. Furthermore, it is accepted that users of social networks, including Twitter, do not reflect the diversity of the community members making use of urban landscapes (Schwartz and Hochman, 2014). Not all members of the urban population engage with smart technologies or social networks, for example older people (75+) show low levels of engagement with these types of technologies (Zickhur and Madden, 2012). While this chapter has shown it is possible to ascertain the gender of a user through a search of their Twitter profile, other demographic information is missing (age, socio-economic status, ethnicity and so on), which are key to a fuller understanding of the cultural demographics operational in forming green space experiences. This limits the utility of Twitter data in creating targeted policies aimed at increasing positive outdoor experiences (Blank, 2016). While the profiling of Twitter users remains a challenge to researchers, advances in the ability to profile an individual based on their online footprint is taking place in other fields; particularly in targeted commercial and political advertising (Xia et al., 2016). Such approaches could be adapted and may offer a solution for social science research in the future.

As a publically accessible and largely unmonitored platform, Twitter and the information displayed in the newsfeeds of users, also have the potential to be manipulated to present false information. For example, fake users can be created which generate spam tweets relating to specific topics. As this study was relatively small it was possible to identify spam accounts and remove their tweets from the sample. However, if adopted more widely and used at larger scales, this important task may become harder, and the presence of spam accounts creating positive or negative tweets may skew the assessment of the subject being studied. Fake accounts which pose as real Twitter users and aim to sway opinion may be harder to identify than spam tweets, especially in large samples.

Methodologically, it should also be noted that the REST API, through which tweets were downloaded for study, relies on an algorithm that Twitter keeps secret (Morstatter et al., 2013). Importantly, it is based on a random 1% sample of all tweets relating to a search query, only making these available to view and download; thus the sample used in this study is by no means comprehensive.

5.6 Conclusion

Tweet text as input for sentiment analysis has proven valuable in ascertaining the positivity/negativity and emotions that are conveyed in short multimedia posts on Twitter about a sample of urban green spaces in Birmingham. It was possible to identify themes of positivity in tweets relating to urban green spaces with happiness being the major positive emotion identified along with appreciation of nature and landscape beauty. Negativity was also identified, albeit to a lesser extent, particularly relating to the emotions of fear and anger. Negative implications of interactions with urban green spaces on well-being have rarely been considered with studies tending to focus on the overwhelmingly positive benefits. Thus while positive responses are more commonly elicited the situation is more complex than previously shown and consideration of the negative responses is important in understanding the barriers to use and enjoyment of these spaces. Using Twitter data it was possible to identify spatial and temporal variation in both the positivity/negativity of tweets in the dataset, as well as in the detected high emotions.

This chapter has demonstrated the potential for using crowd sourced Twitter data in investigations of emotional responses to urban green space. The use of this approach provides a number of benefits to researchers; including the free, publicly available data accessible via the Twitter API and the potential to increase the temporal scale of studies through increased monitoring of study sites. Such information may also be of use to urban planners to support evidence based management and efficient resource use or to monitor responses to specific events occurring within urban green spaces. However, to be utilised most effectively it is important to draw attention to a significant limitation of Twitter data; the lack of demographic information provided about the users creating tweets used for analysis. While in this study the gender of the users was known, there was no information about the age, ethnicity or socio-economic background of the users. It is unlikely that the Twitter demographic is accurately reflective of wider society given that subsections of the population such as the elderly are less prevalent in the Twitter user population as a whole. Thus large subgroups of the population could be missed. Attention must also be drawn to the lack of a control group in this study as the focus was solely on urban green space. In future, examinations of urban grey-space locations would provide a useful an interesting comparison to this dataset and help contextualise the sentiment analysis of both these types of urban spaces.

Chapter 6. Investigating the emotional responses of individuals to urban green space using Twitter data: a critical comparison of three different methods of sentiment analysis[6]

Preface

The possibility of accurately extracting emotions from tweets has been demonstrated in recent studies (e.g. Roberts et al., 2012), which have classified tweets according to a range of readily identifiable and distinct emotions. Chapter Five of this thesis demonstrated that such sentiment analysis can be undertaken successfully in the context of urban green space, making it possible to identify a range of emotional responses to urban green spaces and the events occurring within them from tweet text.

However, working with such an informal text genre presents new challenges for analysts as the language used by the twitter community is often informal with creative punctuation and spelling, slang, abbreviations and URLs (Rosenthal et al., 2014). Debate on how to develop methods which address these challenges and capture the fullest range of responses possible, and how best to mine people's opinions and sentiments is an increasing body of literature.

Rapid developments in automated and algorithm based analysis present new methodological approaches which may help overcome the challenges inherent in using data derived from crowdsourcing and social networks, and have been carried out on Twitter data in a range of contexts (eg. Bollen and Mao, 2011; Bruns and Burgess, 2011; Shalunts et al., 2014; Mitchell

[6] Roberts, H., Resch, B., Sadler, J., Chapman, L. and Petutschnig, A. (2018) Investigating the emotional responses of individuals to urban green space using Twitter data: a critical comparison of three different methods of sentiment analysis, *Urban Planning*, 3 (1)

et al., 2013). Despite the large volumes of work being undertaken in this area, however, there have been few attempts to compare the outputs of different methods of sentiment analysis and to examine how they differ to manual sentiment analysis. Such inter-method comparison and evaluation is critical to establish the relative successes and pitfalls of a methodology but seems to be lacking in much of the current literature. Chapter Six of this thesis begins to address this by presenting the outcomes of three different methods of sentiment analysis undertaken on the same corpus of tweets, in order to compare between the different methodological approaches used.

6.1 Abstract

In urban research, Twitter data have the potential to provide additional information about urban citizens, their activities, mobility patterns and emotion. Extracting the sentiment present in tweets is increasingly recognised as a valuable approach to gathering information on the mood, opinion and emotional responses of individuals in a variety of contexts. This chapter evaluates the potential of deriving emotional responses of individuals while they experience and interact with urban green space. A corpus of over 10,000 tweets relating to 60 urban green spaces in Birmingham, United Kingdom was analysed for positivity, negativity and specific emotions, using manual, semi-automated and automated methods of sentiment analysis and the outputs of each method compared. Similar numbers of tweets were annotated as positive/neutral/negative by all three methods; however, inter-method consistency in tweet assignment between the methods was low. A comparison of all three methods on the same corpus of tweets, using character emojis as an additional quality control, identifies a number of limitations associated with each approach. The results presented have implications for urban planners in terms of the choices available to identify and analyse the sentiment present in tweets, and the importance of choosing the most appropriate method. Future attempts to

develop more reliable and accurate algorithms of sentiment analysis are needed and should focus on semi-automated methods.

6.2 Introduction

6.2.1 Twitter, sentiment analysis and urban green space

Sentiment analysis describes the field of study concerned with analysing the opinions, attitudes and emotions of individuals towards entities such as products, services, organisations, locations and events (Liu, 2012). Over the last two decades, the field has become increasingly active given the vast real-world applications to a plethora of disciplines, such as politics, economics, business, healthcare and urban planning. Increased engagement with sentiment analysis has also coincided with the rapid growth in social networks, without which a lot of the recent research would not have been possible. For the first time in human history researchers have access to huge volumes of freely accessible data published by individuals online.

The increase in social media sites such as Twitter has led to the internet becoming a place of increased expression and opinion sharing on a vast range of topics (Pak and Paroubek, 2010). This phenomenon is providing new sources of text which can be used to gauge public opinion through sentiment analysis (Zhang et al., 2011). Recent studies have indicated the potential and versatility of tweets in examining emotions. These include: a variety of economic (Bollen and Mao, 2011; Jansen et al., 2009; Chung and Lui, 2011) and social (Thelwall et al., 2014) contexts, examining emotional responses to specific events, such as political elections (Tumasjan et al., 2010; Wang et al., 2012; Bruns and Burgess, 2011), natural disasters (Mandel et al., 2012; Shalunts et al., 2014) and terrorism events (Cheong and Lee, 2011); and exploring new ways to measure happiness (Mitchell et al., 2013; Quercia et al., 2012; Dodds et al., 2011). Recent success by Roberts et al. (in prep) identifies how Twitter data can be successfully used to identify both emotions in tweets; and the cause of these emotions, in relation to green space experience. Following the success of this work, this chapter investigates the employment of three different methods of sentiment analysis in this context. In doing so, different methodologies are explored and their limitations discussed.

The information made available by individuals in their tweets has the potential to provide insights into how urban landscapes are perceived by individuals as they navigate them. The urban landscape is being experienced by an increasing number of individuals as global urban populations continue to expand (UN HABITAT, 2016), leading some to question the long-term sustainability of cities (Grimm et al., 2000). Understanding how individuals are responding and relating to city landscapes is a key element for facilitating their design, implementation and management. Urban green spaces in cities provide the opportunity for individuals to have contact with the natural environment (Daniel et al., 2012), a fundamental influence on human well-being (Wilson, 1984; Kellert and Wilson, 1995; Fuller and Gaston, 2009), while the benefits associated with nature and green spaces are a vital component of the ecosystem services they provide to human populations (Costanza et al., 1997; Daily, 1997; Ehrlich and Ehrlich, 1981; MEA, 2005). Despite broad agreement that these cultural ecosystem services are beneficial to urban dwellers (WHO, 2017) there remains limited methodological progress in capturing the transfer and receipt of these services to populations, largely due to their intangible nature and difficulty in assigning economic value to the benefits they provide (Daniel et al., 2012; Milcu et al., 2013). Studies have only recently emerged that consider the effect of number and duration of encounters on ecosystem service receipt (Shanahan et al., 2015), and at present they remain small scale and highly contextualised. Twitter data have the potential to offer a wider spatial and temporal lens through which responses of people to urban green spaces can be captured.

While environmental cues have a significant impact on how individuals respond to and experience space (Ulrich, 1983), a wide range of other factors are also influential, including weather conditions, group dynamics, types of activities and what people observe happening around them. These factors are hard to study successfully due to limitations on experiment size and cohort selection, so capturing their high spatial and temporal variability has proved challenging (Cohen et al., 2009). As a result, studies lack explorations of the emotional responses of people to urban green spaces and the range of sentiments they can elicit in individuals. Twitter data offer the potential to overcome these limitations and can provide information about how individuals feel while experiencing urban green spaces. The information provided in tweets also has the potential to contextualise why an individual may be experiencing certain emotions and what activities they are engaging in that result in the given response. Such information has significant utility for urban planning. For example, data which provides evidence for the beneficial effects of urban green spaces for urban dwellers can be used to justify their continued presence in the urban landscape amidst intense development pressures. Furthermore, the successful identification of the causes of positive and negative emotions experienced by users of urban green space using Twitter data (Roberts et al., in prep), could be used to develop an evidence base from the which planners can create and manage green spaces to promote positive emotional experiences and minimise and remove features which cause negative responses.

Despite the benefits Twitter data can afford to researchers, sentiment analysis studies obtained from tweets are not common, especially in an urban context. Nonetheless, studies have utilised tweet text to investigate how public mood varies both spatially (Bertrand et al., 2013) and temporally (Martinez and Gonzalez, 2013) in urban areas, and to compare how the positivity of Twitter posts by urban citizens varies between different cities (Hollander et al., 2016). Others have used Twitter data alongside additional sources (such as biosensors) to assess how individuals perceive and emotionally respond to cities (Resch et al. 2016), in order to develop more citizen centric approaches to urban planning. For tweets to be a useful source of emotional data to urban planners, methods of sentiment analysis are required which enable the fast, accurate and replicable annotation of tweets.

6.2.2 Methods of sentiment analysis

Numerous works have demonstrated that it is possible to accurately identify and distinct emotions from tweets (e.g. Roberts et al., 2012). However, the inherent challenges associated with working with such an informal text genre present are significant challenges. The language used by the twitter community is often informal with creative punctuation and spelling, slang, abbreviations and URLs (Rosenthal et al., 2014). Analysts are also challenged by the use of emoticons/emojis as the emotions they convey can be highly subjective and often context dependent.

To compensate for the range of challenges inherent in using Twitter data, approaches to identifying sentiment and emotion are varied, but can broadly be placed into three commonplace methodologies. Firstly, manual annotation requires human annotators to categorise tweets into emotion categories (Roberts et al., in prep; Jansen et al., 2009). Fully automated annotation can also be undertaken, relying on algorithms and rules to annotate the emotion in tweets. Many different approaches to fully automated annotation exist, but methods typically rely on n-gram analysis (Barbosa and Feng, 2010) to annotate the emotion in a tweet. Significant limitations have been identified with using both manual and automated sentiment analysis on tweets (and are discussed in detail in subsequent sections). As a result, novel methodologies are being developed to progress tweet sentiment analysis. This study presents one such method, drawing on semi-supervised or machine learning annotation. There are a number of machine learning techniques which can be employed to annotate tweets including Naïve Bayes classification (Pak and Paroubek, 2010; Go et al., 2009), maximum entropy classification (Go et al., 2009), graph based propagation algorithms (Resch et al., 2016) and semantic orientation (Turney 2002). The method presented herein relies on a graph based semi-supervised learning algorithm (Resch et al., 2016) and is described in full in section 2.5. The variety of approaches undertaken within these three methodological approaches reflects the complexity inherent in the task.

This chapter uses tweets relating to urban green spaces to evaluate three different sentiment analysis methods, focusing on the variation in sentiment they indicate, in order to facilitate discussion around the limitations and benefits of each approach. However, this paper does not attempt to identify the most effective method for tweets. Instead, the objectives of this paper are twofold:

1) To compare the outcomes of manual, fully automated and semi-supervised learning methods of sentiment analysis on the same corpus of tweets;
2) To evaluate each method in the context of urban green space research.

The three methods of sentiment analysis presented and compared herein have been chosen as each one is derived from one of the three broad methodologies of sentiment analysis: manual, automated and semi-automated. In this way, a comparison can be made between these differing methodologies in the context of urban green space research; and their potential contribution in providing ways for urban planners to engage meaningfully with social media derived data.

6.3 Methodology

6.3.1 Case study location

The tweets collated for analysis relate to sixty urban green spaces located in Birmingham, United Kingdom (Figure 2.1, Chapter 2). With a population of approximately 1.1 million people (ONS, 2014) the 600 public parks, open spaces and nature reserves within the Birmingham metropolitan area (BCC, 2016a) provide an important resource for urban citizens in terms of their contribution to cultural ecosystem service provision.

The sixty green spaces were chosen to reflect the diversity of spaces found across the city in terms of their size, habitat type, available facilities and amenities and locations within different types of neighbourhoods. Alongside forty six parks, fourteen linear green features were also included for investigation consisting of the footpaths along four rivers and seven canals and three cycle ways.

6.3.2 Tweet corpus creation

The tweets used in this study were obtained via Twitters publically accessible REST API. . The REST API provides access to a 1% sample of tweets published by users with public profiles, and allows queries to be used to search for specific tweets. Searches made using the REST API are based on relevance and therefore this source of tweets was most appropriate for use in this paper. To create the tweet corpus used in this study, English language tweets were downloaded every ten days from the REST API given that tweets made available through the REST API are updated at this frequency. This ensured maximal temporal coverage over a period of twelve months, from June 2015 to May 2016. A search query was used to ensure that the tweets downloaded related to one of the sixty sites included in the study. Therefore, each tweet in the corpus contains specific references to one of the sixty green spaces included in the sample. Any duplicated tweets were removed during pre-processing. In this way, a corpus of 10268 tweets was generated for use in this study.

6.3.3 Manual annotation

During manual annotation, tweets were first assigned into one of three categories: positive, negative or neutral. This annotation was based on the presence of emotive words, emoticons/emojis or meaning. Subsequently, the positive and negative tweets were categorised into distinct emotions. The higher level emotions chosen included five of Ekman's basic emotions (anger, disgust, fear, sadness and happiness (Ekman and Friesen, 1971; Ekman, 1992)), in line with previous research using Twitter data (Roberts et al., 2012; Resch et al., 2016). These emotions are arranged into the ontology shown in Figure 6.1. In this study, beauty was included an additional sub-category to the positive tweets but outside of the emotions to account for the large amount of tweets referencing the beauty of nature and the landscape (as to be expected for green space). Each tweet could only be assigned into one of these emotion categories based on the strongest present emotion.

Five annotators were used to annotate a random sample of 1,000 tweets, in order to ensure there was sufficient agreement between different annotators in how tweets were categorised. A metric of comparison was derived (K=0.666) suggesting sufficient agreement to assume inter-annotator reliability (Landis and Koch, 1977). Given the identification of sufficient inter-annotator reliability between annotators, and the time required for the task, the remaining tweets were annotated by one annotator. To the authors' knowledge this is largest manually annotated dataset of sentiment present in tweets, providing a robust test set against which other methods can be compared.

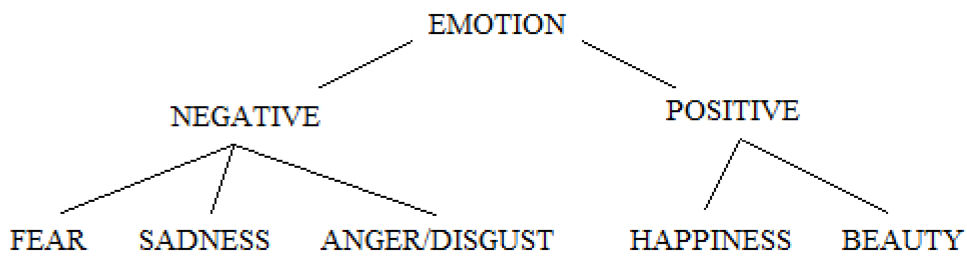


Figure 6.1 *High level emotion ontology for the emotions used in manual and semi-automated tweet annotation.*

6.3.4 Fully automated annotation

For the automated method, an Affective Norms for English Words (ANEW) resource was used as the basis for emotion annotation. The ANEW resource utilised here derives from Warriner et al., (2013) in which over 13,000 English lemmas were assigned valence scores. Using an automated process these scores were used to annotate the valency of each tweet in the corpus. After assigning each word in each tweet with a valence score, an average valence score was created for each tweet based on the number of words present. Thresholds were then used to place the tweets into positive, neutral and negative categories. Following the thresholds used by Warriner et al. (2013) tweets with scores of ≥6.0 were categorised as positive, scores between 5.9 and 4.9 were categorised as neutral and scores of ≤4.9 were categorised as being negative. Given there remains no robust way to determine specific emotions from numeric scores, this method only annotated the tweets in terms of their positivity as opposed to annotating each with a discrete emotion. The implications of this are discussed in greater detail further on.

6.3.5 Graph based semi-supervised learning annotation

In this method (Resch et al. 2016), a sample of manually annotated tweets was used to train a graph based semi-supervised learning algorithm which annotated the remaining tweets. A sample of 1,000 tweets from the corpus, known as the gold standard, were annotated manually (as described in section 6.3.3) and used to train and evaluate the annotation algorithm. This was done to compromise between manual and automated analysis and capture the benefits of each, namely the accuracy of manual annotation and the quickness of automated annotation.

In order to classify tweets according to the emotions they contain a similarity computation was first undertaken, where similarity is defined as the likelihood that two tweets contain the same emotion. The similarity computation comprises three dimensions; linguistic similarity (defined through proven emotion emotion-related linguistic features such as co-occuring words, part-of-speech tags, punctuation, spelling, emojis and n-grams), spatial similarity and temporal similarity (defined through spatial and temporal decay functions according to recent literature). It should be noted that the results presented in this paper only used the linguistic feature groups because not all tweets were geolocated, thus lacked the necessary spatial information.

Once the similarity between tweets has been computed, the graph, which creates the input for the semi-supervised learning approach is constructed and is defined by the tweets (nodes) and pairwise similarity values (weighted edges). Assigning emotions to the tweets was undertaken by applying the graph-based semi-supervised learning algorithm Modified Absorption (MAD) using a subset of the gold standard (training dataset) as this method is found to be most effective for graphs where nodes connect to many other nodes (Talukdar and Pereira, 2010). The assigned emotions were then validated using the rest of the gold standard (test dataset) through computing statistical measures including precision, recall, f-measure and micro average precision. The results prove to be better than random and majority baselines which in the understanding of the field of computational linguistics, demonstrates that the method works well. The developed algorithm outperforming the majority baseline is considered assuring, whereas the better performance compared to random baseline provides strong evidence that the method works well because it demonstrates that the results are not produced by chance, but that significant similarities have been found between pairs of tweets.

Once each tweet had been assigned a discrete emotion using this method, it was then possible to reverse the process and place the tweets into positive, neutral and negative categories using the same ontology as shown in the manual method.

6.3.6 Analysis

Following presentation of the relevant descriptive statistics for each method, various statistical tests were undertaken to assess the significance of any differences in the assignment of the number of positive, neutral and negative tweets by each of the three methods. Fleiss and Cohen Kappa Indexes were then generated to assess inter-method reliability of tweet assignment into each category between the three methods alongside percentage agreement assessments of the three methods in their annotation of each individual tweet.

6.4 Results

6.4.1 Assignment of the tweets into positive, neutral and negative categories

Variation existed in the numbers of tweets assigned to into the ‘positive’, ‘neutral’ and ‘negative’ categories by each of the methods (Figure 6.2). Although for all three methods, the majority of tweets were placed into the ‘neutral’ category, categorisation of tweets into ‘positive’ and ‘negative’ categories showed to be more variable between the three methods (Table 6.1).

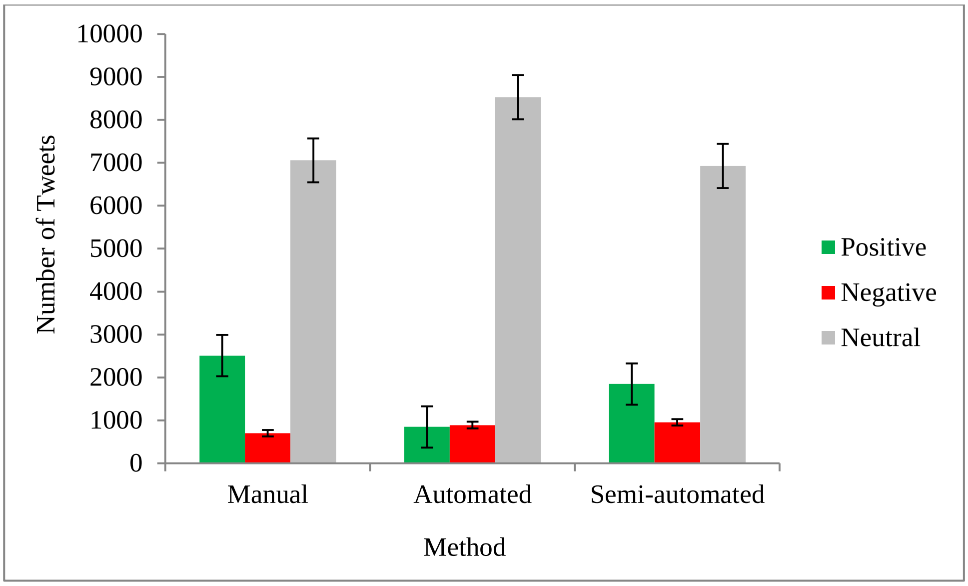


__Figure 6.2__ The number of tweets assigned by each method into positive, neutral and negative categories with standard error bars displayed.

__Table 6.1__ The percentage (%) of tweets assigned by each method to positive, neutral and negative categories.

	Manual	**Automated**	**Semi-automated**
Positive	24.4	8.2	18.2
Neutral	68.8	83.0	68.5
Negative	6.8	8.8	13.3

Given that all three methods show some similarity in the numbers of tweets assigned to each category, statistical analysis was undertaken to investigate the significance of the differences identified between the three methods of classification for all three classes: ‘positive’, ‘negative’ and ‘neutral’. Given that the assumption of homogeneity of variance was not met by the ‘positive’ datasets, a Welch ANOVA test was used and identified significant difference in the number of tweets annotated as positive by each of the three methods ($F(2,18.547)=31.916$, $p<0.001$). Post hoc Tukey analysis identified specific significant differences between manual and automated analysis (p<0.001), manual and semi-automated analysis (p=0.029) and automated and semi-automated analysis (p=0.001). Using a one-way ANOVA, no significant differences were identified between the number of tweets classified as being 'neutral' by each method ($F(2,33)=2.733$, p=0.068). Finally, a Kruskal-Wallis H test, given the violated assumption of normality, identified no significant differences between the number of tweets classified as 'negative' by each of the three methods ($\chi^2(2)=3.466$, p=0.177). Thus, the only significant differences identified between the three methods were in the number of tweets each assigned into the 'positive' category, which was highest for the manual method, followed by the semi-automated and automated methods respectively (Figure 6.2).

By making adjustment to the thresholds (Table 6.2) used to assign the automated tweet scores into the 'positive', 'neutral' and 'negative' categories, it was possible to generate very similar outputs for the manual and fully automated methods (Figure 6.3), and identify no significant differences in the number of tweets each method assigned to each category.

Table 6.2 *Original and adjusted thresholds used to assign automated tweet scores into positive, neutral and negative categories.*

	Original threshold adapted from Warriner et al. (2013)	**Adjusted threshold**
Positive assigned tweets	≥6.0	≥5.73
Neutral assigned tweets	≥5.0	≥4.931
Negative assigned tweets	≤4.99	≤4.93

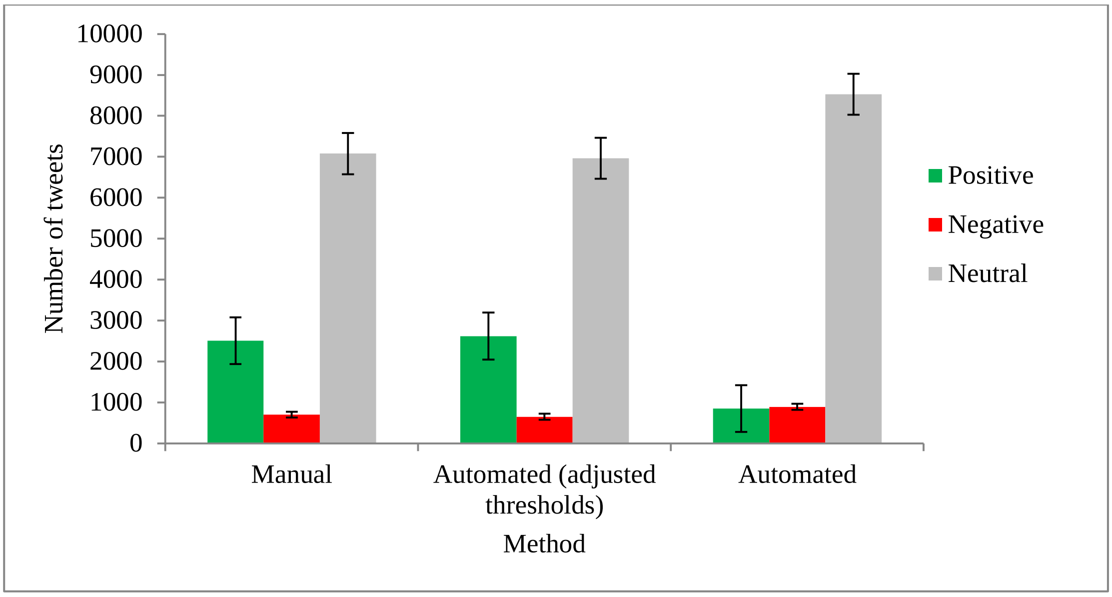


Figure 6.3 *Comparisons of the numbers of tweets assigned to positive, neutral and negative categories by the manual and automated methods using two different thresholds.*

6.4.2 Inter-method reliability

Consideration of inter-method reliability however, shows a more complex picture. A Fleiss Kappa Index identified very little inter-method agreement (k=0.0646) between the three methods, highlighting that the annotation of each individual tweet into the three different categories by each method differed substantially. Indeed, only 47.5% of tweets were found to have been assigned the same category by all three methods, with 5.7% of tweets being assigned different categories by all three methods, indicating wide misallocation.

The relatively high percentage agreement compared to the low Fleiss Kappa Index is due to a large number of tweets being annotated as neutral by all three methods. Indeed, further investigation of the 47.5% of tweets which were annotated the same by all three methods revealed the vast majority to have been assigned to the 'neutral' category (98.6%). However, annotations of positive and negative tweets were less similar, suggesting that where emotions were present, the methods showed more variance in identifying them, either annotating them as neutral or with the incorrect polarity of positivity. Positive and negative annotation agreement between all three methods was extremely low at 1.4% and 0% respectively.

Interestingly, the low percentage in the agreement of tweets remained following the adjustment of the automated thresholds. The adjusted threshold annotations showed most similarity with the manual annotations. Again, however, only 56.8% of tweets were placed in the same category by both methods; showing that despite increasing similarity in number of tweets assigned to each category by each method, altering the thresholds used to assign tweets into 'positive', 'neutral' and 'negative' categories had no effect on increasing the percentage agreement of tweet assignment between the manual and fully automated methods.

Cohen Kappa tests were undertaken to see if the inter-method reliability was higher between any two specified annotation methods. The highest inter-method reliability was found to be between the manual and semi-automated methods (K=0.184), compared to similarity between manual and automated (K=0.0814), and semi-automated and automated methods (K=-0.00961). However, all these Kappa Indices are low (McHugh, 2012) and there remains large variation in the way each method assigns individual tweets into 'positive', 'neutral' or 'negative' categories, despite the appearance of similarity in Figure 6.3.

6.4.3 Quality control using character emojis

By way of a quality control measure, assessment was undertaken on just the tweets containing objective character emojis for the manual and semi-automated methods (automated annotation did not include character emojis in the lexicon). This was done as tweets containing such characters clearly belonged to either the positive or negative categories. All tweets containing positive or negative character emojis were assigned as 'positive' or 'negative' respectively by the manual method indicating a complete success rate of allocating these tweets into the correct emotion category. Compared to this, the ability of the semi-automated method was less successful. 71.4% of tweets containing positive character emojis were misallocated by the semi-automated method as either 'neutral' or 'negative'; while 100% of the tweets containing negative character emojis were misallocated as 'neutral' or 'positive'.

6.4.4 Assignment of tweets into discrete emotion categories

Using the manual and semi-automated methods of annotation it was possible to assign tweets into a number of emotion categories. A comparison of the number of tweets assigned into each of these categories again highlights substantial variation between the methods (Figure 6.4). Both methods showed variation in the number of tweets they identified as belonging to each emotion category. Substantially higher numbers of tweets were annotated as anger/disgust and beauty by the manual method compared to the semi-automated method.

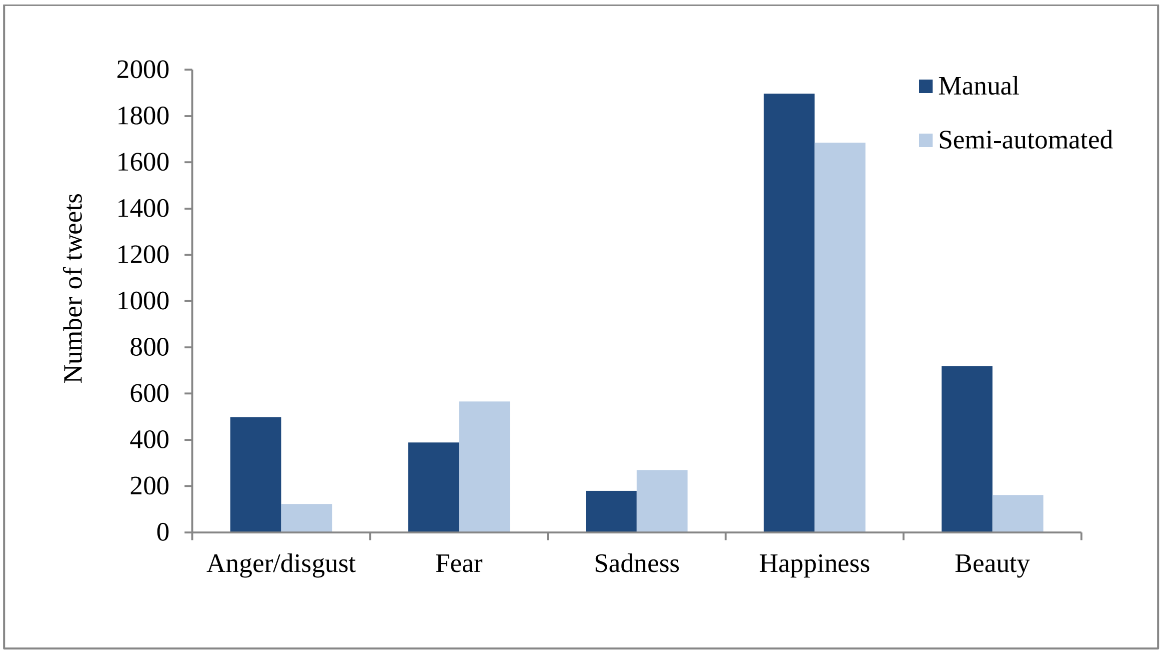


Figure 6.4 *The number of tweets assigned by the manual and semi-automated methods into discrete emotion categories.*

Percentage agreement between the two methods was found to be 47.6% when undertaken on all tweets. However, when tweets which were allocated as 'neutral' by both methods were removed, this figure falls substantially to 4.7%. This indicates that the methods show higher levels of variance when allocating an emotion to a tweet as opposed to just identifying the presence of an emotion, and that the presence of neutrality in a dataset can affect how the results of agreement between the assignment of tweets can first appear. A Cohen's Kappa Index of 0.149 further emphasises the low level of agreement in allocation of tweets to discrete emotions.

6.5 Discussion

6.5.1 Comparison of the outputs of manual, automated and semi-automated analysis

The results presented show that detecting sentiments from tweets is a highly complex task, and importantly, that the method of analysis employed determines the categorisation of positivity, neutrality or negativity, despite using the same corpus of tweets. Moreover, the comparison of the manual and semi-automatic methods illustrated considerable variability in Ekman's specific emotion classes.

All three methods were found to assign variable yet similar numbers of tweets into the positive, neutral and negative categories, with the majority of tweets being annotated as neutral, followed by smaller numbers of positive and negative tweets respectively. Despite this analysis suggesting similarities between the three methods, assessment of inter-method reliability found percentile agreement between the assignment of tweets into the three categories by the two methods to be only 47.5%.

The adjustment of thresholds used to assign automated tweet scores into positive, neutral and negative categories improved the similarity in the number of tweets assigned to each category between the manual and fully automated methods; however, it did not improve the percentage agreement between the two methods.

Manual annotation has previously been cited as providing the most reliable method of sentiment analysis, given that human annotators have the best chance of identifying the emotion present in a tweet (Saif et al., 2013). However, a dataset resulting from manual annotation is not unambiguous given that labelling tweets with an emotion remains a subjective task (Resch et al., 2016). Different human annotators may interpret the same text differently for many reasons – for example, sarcasm, slang or ambiguous use of emojis. This issue is also relevant for the semi-supervised learning method demonstrated in this paper, given that the 'gold standard' tweet dataset used to train the algorithm relied on initial manual annotation of 1,000 tweets. To ensure that annotation was reliable between human annotators, a metric of comparison was derived suggesting agreement between them to be sufficient to assume inter-annotator reliability (Landis and Koch, 1977). Kappa Indexes enable the assessment of inter-annotator reliability between manual annotators and allow the variation in annotation by different annotators to be quantified.

Setting aside inherent subjectivity, the most significant limitation of manual sentiment analysis of tweets is the researcher time needed to examine each tweet. Given that Twitter generates large volumes of tweets in very short time periods, manual annotation is simply not viable. For this reason, automated and semi-automated methods are often employed.

Automated methods of sentiment analysis offer a quick and easy means of annotating large tweet datasets. Methodologically, however, there remains no robust way to derive discrete emotions from numeric scores, thus the granularity of the automated method demonstrated herein is limited to assessment of positivity rather than identifying specific emotions from tweet text. In this study, a large lexicon of words was used to enhance the reliability in the scores generated for each tweet. Despite this, the limitations seem to outweigh the benefits. Low inter-method reliability was prevalent and there was a particularly low percentage agreement between annotations of positive and negative suggesting that this method is unlikely to reliably identify the correct polarity of sentiment in tweet text. Additionally, while the large lexicon used provides robustness for scoring words, it does not include emojis which are increasingly common ways to express sentiment in short social media posts (Pavalanathan and Eisenstein, 2015). Previous research has shown that emojis can be successfully used to inform automated analysis of tweets (Go et al., 2009). Indeed, the creation of an emoji lexicon in which each is given a score would be of significant use to future research and enable the combined use of words and emojis in the annotation of sentiment from tweet text. Such an undertaken would need to overcome the challenge of interpreting emojis in their different representational forms: Unicode (e.g. “U+1F642”), Kaomojis (e.g. “(◕‿◕)”), a sequence of ASCII characters (e.g. “:-)”) or a specific code used by Twitter (e.g., “<ed><a0><bd><ed><b2><af>” or “<ed><U+00A0><U+00BC><ed><U+00BC><U+009E>”).

An issue of spatial variation in language use was also identified associated with the automated method of annotation. Despite the large lexicon used, it cannot account for regional/local dialect. Given the location for this study was Birmingham, where some language used by local populations is not used elsewhere, these words will not have been included and scored and a proportion of sentiment in the tweets, albeit small, will not have been captured by this method. Provided that manual annotators are native to the language and region from which the tweets have been captured, this should not be an insurmountable issue.

The semi-automated method generated similar numbers of neutral, positive and negative tweets as the other two methods. However, Kappa Indices indicate that the placement of individual tweets into each of these categories showed low levels of agreement. Differences were also identified in how semi-automated annotation assigned tweets to discrete emotion categories, when compared to manual annotation. The notion of beauty is not a basic emotion as defined in emotion psychology; indeed, it is usually subsumed under happiness. This makes it difficult for the algorithm to identify beauty in written text because it is often expressed in comparatively subtle terms.

For the experiment presented in this paper, it was possible to identify a limitation in the semi-supervised method, in that the full range of emojis in the dataset could not be captured by the algorithm. The method is designed for character-wise emojis (e.g. ":-)"), however unicode emojis are widely used alongside character-wise emojis in tweet texts. In fact, the semi-supervised learning method was not able to interpret unicode emojis, increasing the likelihood that essential elements of tweets were missed by this method, diluting the precision of assigning emotions and polarities.

The quality control measure, which used character emojis to assess the allocation of tweets into the correct category, highlighted that the semi-automated method was often unable to recognise emotion, despite these being included in the assessment of linguistic similarity undertaken during analysis.

The parameter choices of semi-automated approaches make such methods highly sensitive; the number of seeds used, the seed distribution, details of similarity computation, edge weight threshold and the emotion categories used strongly influence the results. A significant issue is that no formalised method exists to perform an *a priori* estimation for these parameters. In most cases, 'optimal' parameter settings can only be found through empirical experiments, which in turn means it cannot be stated with certainty how good any results are in relation to the best achievable results. Thus, the parameter choices require a substantial amount of expert knowledge and experience, particularly because random permutations cannot be performed due to the computational complexity of the algorithms. This opens up debate as to how a training dataset should be generated. In this paper, 1,000 tweets were randomly chosen. It may be more appropriate to actively identify tweets which cover all the discrete emotion categories so the algorithm can learn most effectively.

Finally, in this paper, for all the methods of emotion annotation used, it was assumed that one tweet contains a maximum of one emotion. However, in reality tweets can be inherently more complex and contain a variety of emotions over a short space of characters. This is a finding that future methods looking to classify the emotion in tweet text will need to consider and overcome to provide the most accurate interpretation of the emotional information that tweets contain.

6.5.2 Implications of these findings for urban planners

The availability of emotional data to urban planners has significant utility in the creation, management and justification of urban green spaces which promote positive emotional experiences and minimise features which may elicit negative emotional responses (Roberts et al., in prep). The provision of such emotional data through social networks, such as Twitter, provides the opportunity for planners to gain access to this information in inexpensive, time efficient and replicable ways. However, in order to be used meaningful, methodologies are required which can accurately annotate any emotion present in a tweet relating to an urban green space.

This article has identified that challenges remain to this end. Indeed, none of the three methods presented herein are appropriate in their current form to provide sentiment analysis of tweet text for urban planners. Whilst manual analysis can be used to accurately identify any emotion present, the amount of time taken to undertake this method on a large corpus of tweets makes it unsuitable in the context of urban planning where resources and individuals are often limited.

Similarly, the current inability of automated and semi-automated methods to accurately identify emotion, make them dubious approaches to employ where the identification of such emotion and their causes could have significant implications for the management and creation of green spaces.

However, the authors tentatively suggest that pursuing a semi-automated method, like the one presented herein is the most appropriate way forward. The development of a method through which the accuracy of manual annotation can be achieved, in much shorter time is doubtless of interest to urban planners. This is of particular relevance because manual annotation of tweets is a time-consuming and expensive method. This article suggests that the development of a gold standard training data set should be a priority, enabling algorithms to learn the variety and complexity with which emotions can be conveyed in tweets.

Without a doubt, Twitter data presents a useful and abundant source of easily accessible emotion information which is generated by users as they experience specific urban green spaces. Such a source of data presents vast opportunities for urban planners; however there remains a need for increased innovation and development in the methodologies which would enable this data source to be engaged with most effectively.

6.6 Conclusion

This paper has presented a comparison of three approaches to sentiment analysis undertaken to collate the sentiment and emotion present in tweet text. Despite their utility, this paper has identified significant differences in the outcomes of three methods of sentiment analysis on the same corpus of tweets, and the discrepancies in how tweet text is analysed by different methods is a critical consideration for future research.

It was possible to identify differences in positivity annotation between all three methods in terms of the numbers of tweets they assigned to each category as well as inter-method reliability in assignment. Using the manual and semi-automated methods, discrete emotions can be annotated, but again significant differences were identified in this process, particularly for beauty and anger/disgust tweets.

Overall, whilst this paper is positive about the role of Twitter in providing a useful and substantial data source for urban planners on which to undertake sentiment analysis, it suggests caution is needed in interpreting the outputs of sentiment analysis and an understanding of the process can help place the results in an appropriate context. A critical discussion of the limitations identified through the undertaking of all three methods in this research has been presented. In doing so, this paper adds to the debate surrounding annotation of sentiment and emotion from tweets and identifies methodological constraints which should be taken into account in future work. Given the utility of the sentiment information captured by tweets relating to urban green space for planners and decision makers, it is of important that an efficient and reliable method is established through which these can be identified and annotated. Despite its reliability, manual annotation remains unfeasible for use on large volumes of data. However, automated and semi-automated methods remain hampered by a number of limitations associated with each, and this paper suggests methodological progression is necessary before either can be used robustly to annotate the sentiment from large tweet datasets.

The findings presented in this paper suggest that automated methods of sentiment analysis are not able to accurately identify the emotion present in tweet text and that manual analysis, whilst accurate is impractical for use on large tweet corpi given the time taken to undertake such analysis. As a result, this paper suggests that future attempts to develop methods of sentiment analysis should focus on semi-automated methods, with particular focus given to how the gold standard dataset is created. Successful algorithms should be aim to include Unicode as well as character emojis in order to best capture the emotion represented by these in tweets.

Chapter 7. Using crowdsourced geospatial data to explore the dynamic nature of human interactions with urban green spaces: an emerging opportunity?

Preface

The previous chapters presented in this thesis provide demonstrations of how Twitter data can be used successfully to monitor interactions between urban individuals as they record their activity in an urban green space. These recordings, due to the nature of a tweet, are static in both space and time and are isolated from the wider context of the journey from which they are a snapshot. This chapter theorises that for social media and app generated data to progress investigation into urban socio-ecological interactions, there is a need for them to be able to capture the spatio-temporal dynamic nature of these relationships.

As functional locations, urban green spaces provide opportunity for human engagement with beneficial behaviours such as physical activity (Bjork et al., 2008; Niemela et al., 2010; Lee and Maheswaran, 2011), active recreation and engagement with nature (Maller et al., 2006); significant cultural ecosystem services. Engagement with these activities is associated with benefits in both physical and mental health (Warburton et al., 2006; Penedo and Dahn, 2005). Beyond these cultural ecosystem services urban green spaces offer a plethora of other ecosystem services (Costanza et al., 1997), such as their ability to mitigate against noise and air pollution and aid microclimate regulation, which are of critical importance and impact substantially on the quality of life of urban inhabitants (Bolund and Hunhammar, 1999).

In order to understand how these ecosystem services are delivered to the population, knowledge of how green spaces within the urban matrix are utilised by urban citizens is urgently needed. In essence, it requires understanding how people interact with green space rather than just identifying their location. Interactions between people and urban green spaces are not discrete events in time in one location; rather they occur within a wider spatial and temporal framework. Indeed, people's interactions with urban green spaces can be considered as dynamic events in space and time, and the dynamic nature of people's interactions with urban green space remains poorly understood. This chapter explores the potential of data from Twitter, BetterPoints and other apps and social networks in facilitating more nuanced understandings of human mobility in an urban context.

7.1 Abstract

A considerable body of evidence now exists highlighting the value of urban green spaces as providers of numerous beneficial ecosystem services in urban areas. The spatial configuration of green spaces and the efficacy of the ecological services they produce has been estimated based on the location and structure of different types of habitats in the urban landscape. This information alone, however, is inadequate to estimate the ecosystem services that urban dwellers receive. Static identification takes no account of the daily movement of people throughout a city, whether commuting to work or school, visiting friends or going to the shops. The movement of an individual through the urban landscape and the routes chosen will impact on the interactions they have with urban green spaces and the ecosystem services they may or may not receive as a result. This chapter presents two case studies to assess whether new sources of crowdsourced data from apps and social networks may be useful in furthering the ability of researcher to investigate the dynamic relationship between urban individuals and the urban green spaces they pass through, or near as they navigate the cityscape. Although preliminary, the results suggest that data obtained from apps and social networks provides

opportunities for urban research, with the results of these investigations having practical application in urban planning and decision making.

7.2 Introduction

Globally, an increasing number of people are living in urban areas (UNHABITAT, 2016). In England and Wales, the 2011 census found 81.5% of the population to be living in urbanised environments (ONS, 2013). These landscapes offer human populations fewer opportunities for interactions with nature and ecosystems. Therefore, green spaces in the urban matrix are of considerable importance for city dwellers in terms of the health benefits they can afford. Indeed, there is a recognised need to better understand how social factors interact with urban landscapes and ecosystems to produce socio-ecological dynamics (Grimm et al., 2000; Redman et al., 2004; Yli-Pelkonen and Niemela, 2005; Ernstson et al., 2008). Where present, natural ecosystems make important contributions to public health through the provision of ecosystem services to urban populations (Barbosa et al., 2007).

The increased prominence of a discourse surrounding ecosystem services in the literature reflects the understanding that human health and the sustained wellbeing of both individuals and populations is intrinsically linked to the natural environment (Gomez-Baggethun and Barton, 2013). There are numerous determinants on human health which are often complex and interrelated (Galea and Vlahov, 2005). Environmental conditions are a key factor, among others, which have significant implications for human health (Whitehead and Dahlgren, 1991).

Recent arguments have emerged to suggest that geography largely works to apply spatial concepts as static notions (Dijst, 2013). To progress, research needs to account for how ecosystem services themselves are dynamic within cities, and move away from traditional site bounded assessments of ecosystem service provision based purely on assessing ecosystem service provision from the structural function of the environment (Syrbe and Walz, 2012). Indeed, ecosystems and their beneficiaries are not always co-located and ecosystem services cannot be depicted as static phenomenon (Fisher et al., 2009). However, Tallis et al., (2008) identify that investigation into ecosystem services has not yet begun to consider the importance of human movement in governing the dynamics of ecosystem service provision.

The flow perspective on cities states that no component of the urban system is static, indeed locations in space are interconnected through flows of, for example, people, goods, energy, information, waste, water and air (Cresswell, 2004). Without quantifying actual flows and use of services, their true value for the benefit of urban populations cannot be understood (Bagstad et al., 2013). Mapping of ecosystem services flows remains challenging (Tallis and Polasky, 2009) as true assessments of ecosystem service provision require consideration on both human and natural systems and the complex interactions which occur between the two (Johnson et al., 2012).

Human mobility takes the form of numerous space-time paths drawn by individuals as they navigate this dynamic landscape (Dijst, 2009). As an individual moves through the urban landscape, the path they choose influences the ecosystem service benefits with which they are exposed. However, quantifying ecosystem services is extremely challenging and numerous discussions are ongoing in the literature as to the most appropriate way to capture them (Plieninger et al., 2013; De Groot, 2010; Sherrouse et al., 2011; Bagstad et al., 2013). This coupled with the difficulties in gathering movement data on the millions of people that navigate urban landscapes on a daily basis, make studying the receipt of urban ecosystem services to dynamic individuals extremely challenging. As a result, such investigation remains limited and there is a need to develop methods which consider the dynamic processes inherent in the relationships between humans and ecosystems, in order to expand the capacity to address fundamental questions about socio-ecological systems (Carpenter et al., 2009). It is beyond the scope of this thesis to discuss the best way to quantify ecosystem services and how their delivery to human populations can best be captured. However, this thesis posits that if methods can be developed to capture human movement successfully at scale, then the problem of investigating dynamic ecosystem receipt can begin to be addressed. Ecosystem service flow traditionally defines the transmission of a service from ecosystem to people (Bagstad et al., 2013). Understanding how people move is one part of the equation for understanding the flow of ecosystem services. It is necessary to know how and where people are moving through the landscape to make investigations into how and where people are receiving ecosystem services possible.

Technological developments in the ubiquitous form of smart phones and hand held devices, and the increased connectivity of individuals with others and larger infrastructures present such a possibility. Smart phones are ubiquitous among urban citizens in the United Kingdom and their GPS functionality enables the tracking of their owner's position in the urban landscape. Thus, vast amounts of data are being generated constantly by millions of individuals as they move.

Initial investigations into using technologically derived datasets to map human movement in urban landscapes relied on location based service information from mobile phones. For example, Ratti et al. (2006) mapped mobile phone usage to create graphic representations of the intensity of urban activities and their evolution through time and space. Similarly Van der Spek et al. (2009) used GPS data to track the movement of a small cohort of pedestrians through city centres in Norwich, Rouen and Koblenz.

With the development of social networks and other smart phone apps in the late 2000s, new sources of data have been made available. Indeed, studies have shown the utility of such data in mapping human movement in urban areas drawing on a variety of sources including Foursquare (Silva et al., 2014; Phithakkitnukoon, S. & Oliver, 2011), Instagram (Silva et al., 2013), Flickr (Sun et al., 2013; Clements et al., 2010) and Strava (Musakwa and Selala, 2016; Sun, 2017). However, this information has yet to be employed in the context of urban ecosystem research and used to gain insight into the delivery, exposure and receipt of ecosystem services to and by urban individuals.

The structure of this chapter differs to those that precede it; no explicit results or analyses sections are provided. Rather, this chapter presents an exploration of the extent to which social media and app generated data may be useful as a source of information for research into dynamic relationships between people and urban green space, and how this affects the ecosystem services an individual receives as they navigate urban landscapes. In short, can these data be used to progress investigations of ecosystem service delivery and receipt from the static to the dynamic by capturing the movement of individuals as they navigate urban landscapes?

7.3 Datasets

As with the previous chapters presented in this thesis, this chapter uses the metropolitan area of Birmingham as the location for its investigations. In Birmingham, publically accessible green areas account for 24.6% of urban space (Guardian, 2017) in the metropolitan area (Figure 1.2, Chapter One), making them a significant component of the urban matrix with which people must pass through or near as they navigate the city.

The datasets utilised to illustrate the discussion points in this chapter have been provided by two applications available on smart devices: Twitter, a popular social networking site, and BetterPoints, a locally available physical activity tracker and reward app. Each collects data on their user movements through the location based services inherent in smartphone technologies. GIS has already been highlighted as an effective environment in which to implement time-geographic constructs (Kwan, 2004) making it suitable as a source of data in this chapter. The increasing availability of geo-referenced data depicting individual movement and the increasing capability of geo-computational methods is creating opportunities for investigating human activity patterns in a meaningful way (Kwan and Lee, 2003). Smart devices and apps afford the ability to digitise human traces, and provide vast opportunities for developing new methodological approaches (DeLyser and Sui, 2012) to answer a plethora of research questions.

Twitter has thus far provided the dominant source of data used in this thesis. In this chapter, a dataset of 56,169 geolocated tweets has been kindly provided for analysis by Dr Wendy Guan from Harvard University's Center for Geographic Analysis. This dataset is comprised of all geolocated tweets created by Twitter users in Birmingham between September 2012 and April 2014 which fall within the green spaces of the city.

The second dataset used in this chapter has been provided by BetterPoints Ltd. Founded in 2010 and describing themselves as an 'evidence-led sustainability, health and social behaviour change technology company', BetterPoints Ltd aims to provide motivation for individuals to engage with more sustainable and social behaviours. Through their smartphone app interface a user earns points for undertaking a number of activities, including walking, cycling and running. The points earned by an individual translate to a financial reward which can be exchanged for shopping vouchers or donated to a chosen charity. Recognising the power of self-tracking in encouraging positive behavioural change (Choe et al., 2014), the app aims to motivate people to make healthier choices for small rewards as a way to promote long term behavioural change in an individual. BetterPoints Ltd are currently active in Hounslow, Hackney, Reading, Sheffield and Birmingham with individuals in these locations being able to sign up to the app for free, log their activities and receive financial rewards. This study makes use of data from the BetterPoints app for the Birmingham area over a twelve month period (June 2015 to May 2016). Over this period users could log their walking, running and cycling activities through the app. These logged activities are stored on a central database and available to the company for analysis. BetterPoints have kindly given permission for the use of walking, running and cycling data for the period of study.

7.4 Results and Discussion

7.4.1 Spatial and temporal variation in urban green space use

The process of mobility is spatially and temporally dynamic and therefore difficult to measure (Hanson, 2005), but social media and app data could provide a method and data source in the geographers arsenal to aid investigation of this phenomenon. Plotting the geolocation of a sample of tweets reveals their spatiality within the urban green spaces in Birmingham (Figure 7.1).

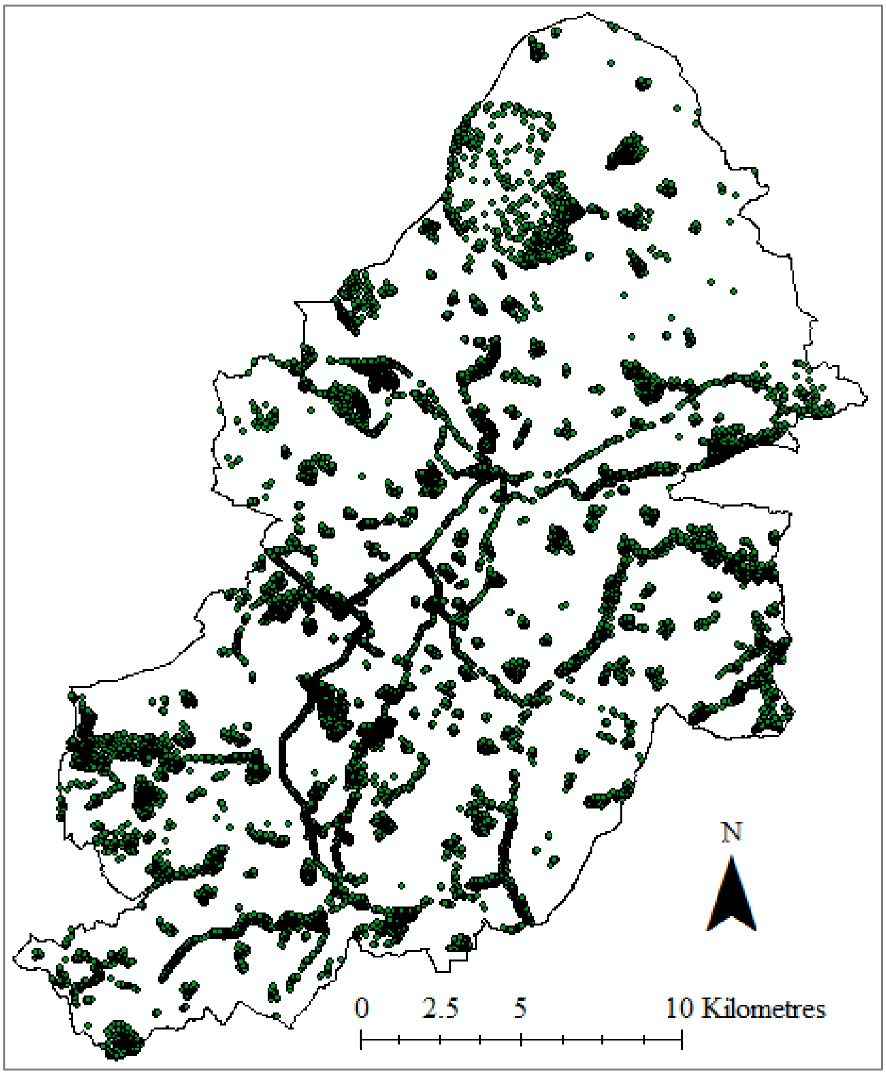


Figure 7.1 *The positioning of the geolocated tweets in the sample.*

While this provides little information beyond where users tweet, density analysis can be used to identify popular locations based on the number of tweets being created in space, and create more nuanced understandings of the temporal variation in urban green space use (Figure 7.2). Identifying spatial hotspots of activity is useful for urban planners, particularly for managing tourist locations (Garcia-Palomares et al, 2015) or quantifying the attractiveness of place (Girardin et al., 2009).

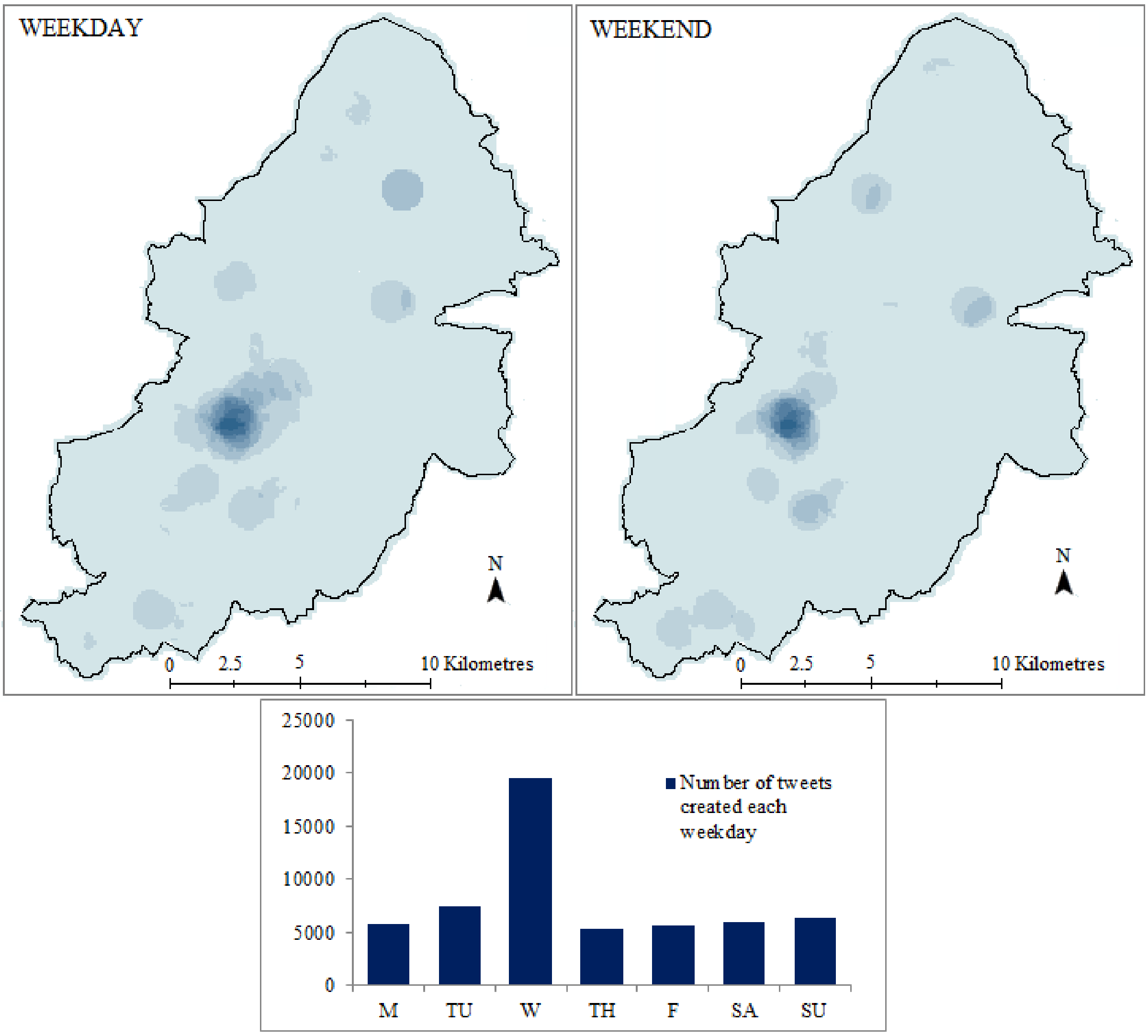


Figure 7.2 *Temporal variation in urban green space use as identified using Twitter geolocation.*

Similarly, density mapping (Figure 7.3a) and subsequent optimised hotspot analysis (Figure 7.3b) reveals a number of spatial patterns in the BetterPoints activity data. For all the recorded activities a large area of disconnect was identified between two major areas of significant hotspot activity. This was particularly prominent for walking and to a lesser extent running, where a number of large coldspots were identified. A clear distinction between two hotspots of activity in the north and south of the city were identified, present across walking, cycling and running. This was also present, to a lesser extent in the hotspot analysis of green space tweets. Such information indicates a disconnect between the north and south of the city and a lack of appropriate route ways for people to use in these areas, and therefore be of use to urban planners informing where connectivity could be increased in the city. In short, both datasets identify spatial patterning and enable the identification of popular tweeting locations within green spaces, and the main route ways that BetterPoints users take giving an indication of movement through the city. Birmingham City Council has identified the importance of making more of the green infrastructure in the city, including parks and canal sides, to encourage physical activity in places other than traditional venues like leisure centres (BCC, 2016b). The kind of data presented here would make it possible to identify where people already use green space for various physical activities, as well as providing information on locations of disconnect which may benefit from investment and improvements in the network.

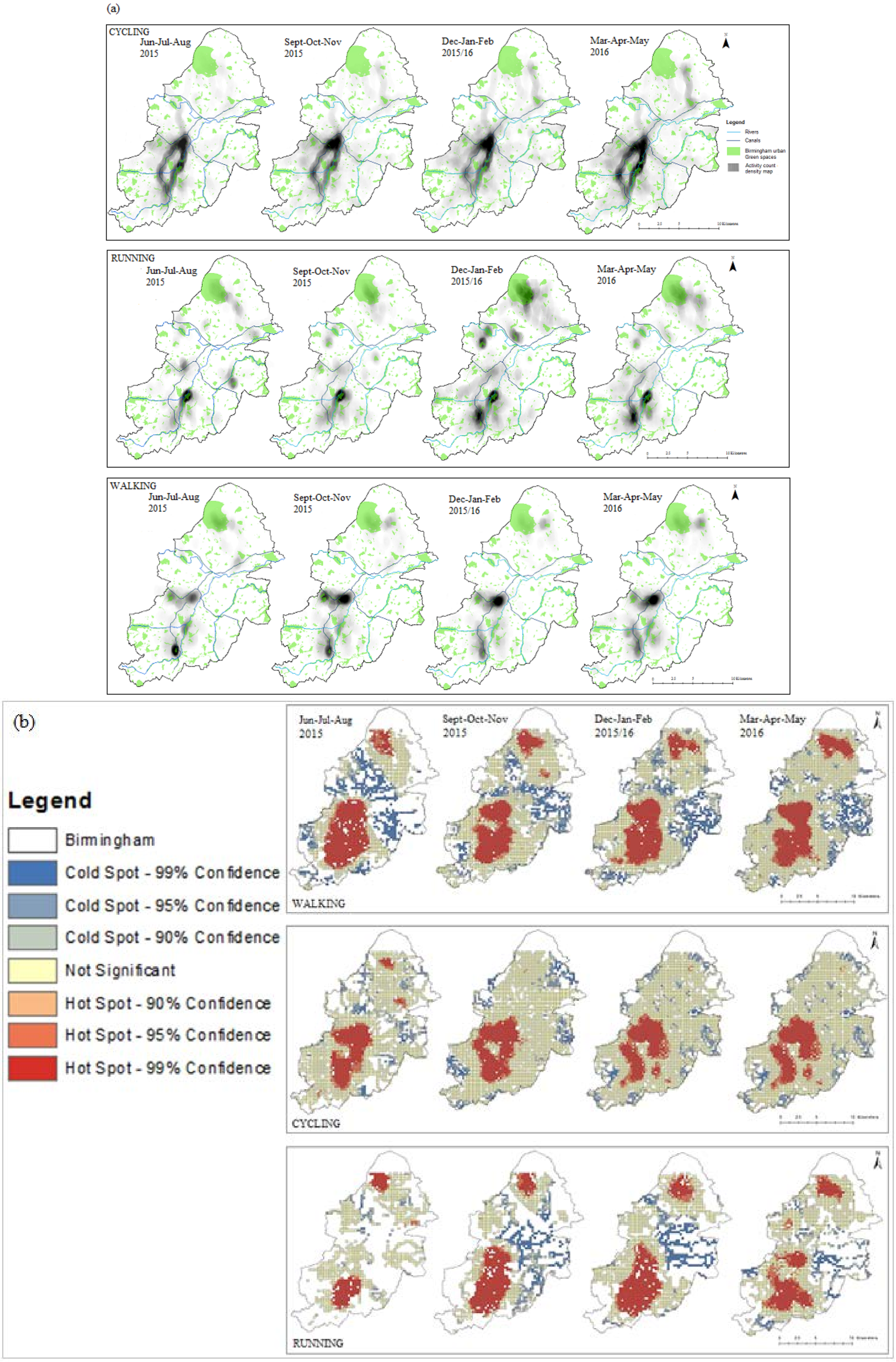


Figure 7.3 (a) *Density mapping of the geolocated BetterPoints data shown with urban green spaces in Birmingham.* ***(b)*** *Optimised hotspot analysis of the geolocated BetterPoints data, with red and blue indicating significant hot and cold spots of activity respectively.*

Given that density analysis identifies the most frequented areas, these data also provides an indication of the accessibility of an urban green space, and can be used to identify areas which have limited use reported use, perhaps due to a lack of public access. Such information could inform the planning action priority to improve the accessibility of sites (Handley et al. 2003), identifying spaces in need of footpath maintenance or additional access points, as well as creating opportunities to create dialogue between the city and its citizens. As shown in Chapter Five, tweet textual information could also be analysed to identify access inhibitors, such as litter and vandalism, enabling the implementation of informed green space improvement initiatives.

Social networks such as Twitter have vast user bases. This, coupled with the ten metre position resolution of tweet data points, affords the opportunity for studying human mobility dynamics at both high resolution and at large spatial scales (Jurdak et al., 2015). Similarly, the resolution of BetterPoints data make it a source of high spatial resolution data. Kabisch et al. (2015) discuss the notion that cities are made up of flows, and thus the spatial scales of research can limit understanding the true importance of ecosystem service provision. To ensure scale does not limit the applicability and interpretation of relationships between urban green spaces and human individuals, a non-static approach using data at a high spatial resolution is required. This chapter posits that apps and social network data offer such an opportunity to researchers. In this application, the high resolution of BetterPoints data ensures greater detail to planners than just an identification of the green spaces people choose to use to move through the city. It enables the identification of exact route ways which are popular within these green spaces (Figure 7.4), and identifies the linear features popular with users of all three activity types. This goes beyond generic awareness of popular spaces, and enables precise identification of the exact parts of a space which are most often used.

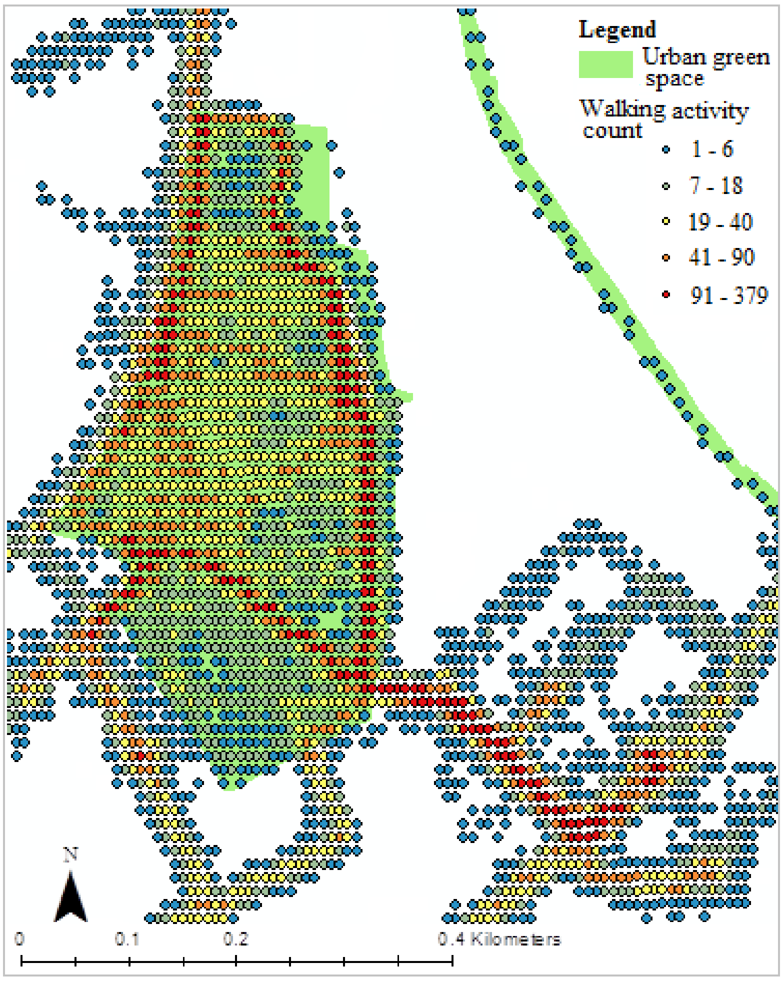


Figure 7.4 *An example of the utility of high spatial resolution BetterPoints data in identifying popular walking route ways through Cotteridge Park, Birmingham.*

7.4.2 Assessment of ecosystem service provision

The locations that urban green spaces provide for physical activity engagement is a key cultural ecosystem service (Han et al., 2013; Bedimo-Rung et al., 2005; Sallis et al., 1998) and the GPS technology inherent to smart devices is a promising tool for understanding the spatial context of physical activity (Krenn et al., 2011). A number of apps, including BetterPoints, enable users to track a walking, running or cycling route and then upload and share this information with other users (Figure 7.5).

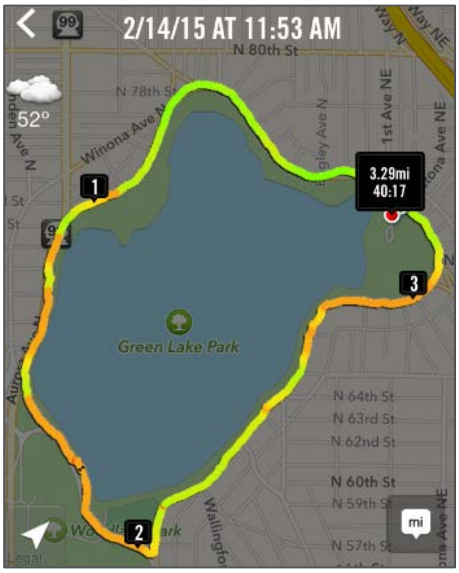


Figure 7.5 *An example of a route taken by a runner recorded using the Nike+ app.*

To do this the user takes their smart device along the selected route and the GPS functionality records their movement. There are a number of apps which provide such a function, enabling users to monitor their physical activity including Strava, Endomondo, Mapmywalk, Mapmyrun, Mapmyfitness, Mapmyhike, Mapmyride, Everytrail, Nike + running, Runkeeper, iBike and Cyclemeter. The data made available via these apps offers presents opportunity for a variety of research endeavours, particularly in relation to ecosystem service exposure. Visualising and digitising the routes taken through the urban environment could provide insight to urban planners as to the types of ecosystem services which are experienced by an individual, and aid assessment of individual exposure to various environmental characteristics, such a pollutants, over daily life schedules (Chaix et al., 2013).

One aspect of some physical activity recording apps of benefit to researcher is the distinctions they make between different user groups. For example, Strava allows users to designate a ride as a “commute”, presenting opportunities for urban planners and researcher to gain information on a specific subgroup of the urban population who engage with sustainable transport behaviours on a regular basis. The identification of popular commuting routes may enable planners to make targeted improvements to these route ways, for example through greening, implementing traffic calming measures (Broach et al., 2012) or adding bike lanes (Krenn et al., 2014); all of which have been found to increase the likelihood of use by running, walking and cycling commuters. Knowledge of the preferred routes of certain subgroups is useful as the choice of route taken by a commuter each day affects their exposure to environmental risks, such as pollution (Zuurbier et al., 2010) and ecosystem services, such cooling provided by shading. The identification of popular commuting routes may enable more accurate assessment and quantification of exposure to these environmental attributes and presents opportunities for app development, allowing users to select routes to minimise or maximise service provision.

The potential of using physical activity recording apps to trace and identify popular activity routes is also significant. Data recording actual use of green space is useful to inform urban planning (Farkas, 2014; Le Dantec, et al., 2015) as once a space has been identified as a valuable physical activity resource it can be maintained as such, continuing to facilitate physical activity engagement among urban populations (Frank and Kavage, 2009).

7.4.3 Theorisations of urban mobilities

Crowdsourced data showing the routes taken by individuals as they navigate urban space have utility in theoretical debates around how relationships between people and space are conceptualised. For example, time-geography theory (TGT) presents a useful theoretical framework through which to visualise dynamic interactions between humans and urban green space within the city-scape, through space and time. Traditionally it has provided a transdisciplinary framework in which space and time provide dimensions of analysis for dynamic processes carried out by an individual (Lenntrop, 1999; Kwan and Lee, 2003). As a flexible way of examining human interactions with space (Thrift and Pred, 1981), TGT provides a suitable framework for investigations into the dynamic interactions between humans and the landscape in an urban context and has already been used as the theoretical framework for numerous works in transport planning (Geurs and Van Wee, 2004; Vandenbulcke et al., 2009), female mobility (England, 1993; Friberg, 1993; Dyck, 1990) and migration (Odland, 1998). A key critique of TGT is that it fails to capture how individuals move and behave in response to the landscape they find themselves in. Spatially located, crowdsourced route data enables exploration into human agency and the extent to which the environment can facilitate and encourage types of human behaviour. In the context of urban research, for example, questions surrounding the influence of green linear features as facilitators of sustainable commuting behaviours such as walking or cycling could be explored; and whether a lack of such features acts to constrain this type of behaviour in urban dwellers. Linked to this, ideas of physical environments as 'taskscapes' (Ingold, 2000) and the extent to which urban green spaces act as enablers of movement for urban dwellers (Howe and Morris, 2009), could be explored. Such debate has implications for urban planning and how spaces can be managed to encourage certain goal-oriented behaviour. If urban green spaces act as enablers for physical activity, it is important to identify the features of these spaces that encourage this behaviour; ensuring those features are replicated in other locations.

Additionally, crowdsourced data are well placed to facilitate post-disciplinary theorisations of mobilities. The identification of a 'mobility turn' (Urry, 2007), representing and increasing engagement between the social and physical sciences in analysis and conceptualisations of mobility, requires data to progress analyses which have been "historically static, fixed and concerned with predominantly a spatial social structure" (Urry, 2007, p. 6). Crowdsourced data are suitable to inform such approaches, particularly in an urban context where human movement is increasingly constant, complex and varied, requiring recordings that are able to reflect the intensity and vastness of movement (Hasan, 2013).

There is also the potential for smart GPS tracking to improve the efficiency of urban mobility schemes such as bicycle-sharing, and to understand how these systems are used by urban dwellers. Smart device apps could be used to provide real-time information as to the whereabouts and availability of bicycles for urban dwellers. This utility has been demonstrated in Germany with the 'Call a Bike' scheme which operates in a number of cities (Midgley, 2011).

7.4.4 Methodological Implications

Methodologically, the ability to access crowdsourced data generated via apps and social networks, in which subjects are not aware of being monitored, provides an additional benefit to researchers as a non-obtrusive method of participation. This minimises the risk of participants altering their behaviour as they are aware they are under study, providing the opportunity to monitor park use as accurately as possible compared to previous studies which have relied on report-based, qualitative methods as demonstrated by Sotoudehnia and Comber (2011) in their investigation of peoples perceived access to urban green spaces in Leicester, UK.

Additionally, individually owned smart devices are ideal for low cost route data collection as the recording device belongs to the individual and there is no cost to the researcher to buy and distribute the technology among participants. Users are also accustomed to carrying their devices so data can be collected with minimal effort from the participant.

While the rise in location enabled devices has expanded opportunities for personal location and tracking data collection in urban areas, advances in technology also bring challenges to researchers, presenting a requirement to balance the potential to collect almost unlimited amounts of data with the need to reduce participant burden and develop a system useful to a variety of stakeholders (Cottrill et al., 2013). The vast datasets made available to researchers through apps and social networks requires significant computational power for appropriate analyses to be carried out.

Other issues in user confidentiality and privacy may also arise if research utilises the provision of individual route data, compared to density mapping. For commute journeys in particular, the ability to identify a user’s home and place of work will need to be addressed, especially when users upload their data to public sharing platforms, without full knowledge of the research their journeys may end up informing.

As identified in previous chapters, social network user populations are not without their biases, and with little demographic information provided about their users there remains limitations in the utility of these data for specific research questions, particularly based around social and cultural investigation. Fitness and physical activity based apps, despite their utility as a new source of data, also have the inherent problem of data bias. They are used by people to monitor physical activity and as such exclude members of the population who either do not engage in physical activity, or do not monitor it in this way. They also limit their user base to a subset of the population; those who use their phones online, those who are happy locating themselves with GPS and those who are comfortable taking their mobile device with them whilst being physically active. Previous study has found significant skewness of the user population towards males, predominantly aged between 25 and 54 years (Griffin and Jiao, 2015). Gender biases such as these will have implications for the utility of these datasets in explorations of women’s mobility and make them less suited to investigating movement among less represented groups. However, they can provide a much larger sample population than usual in traditional survey studies.

7.5 Conclusion

This chapter posits that new sources of geolocated data generated by social networks and apps present opportunities for exploring how individuals move through urban landscapes and the routes they choose to take. Such information is an important first step to better understand interactions between people and urban space, and how the spatially and temporally dynamic nature of this relationship impacts upon the ecosystem services an individual may receive as they move through the city. Such a notion extends current research methodologies in which the localities of ecosystem services provision can be roughly known based on the location of different types of green spaces throughout the urban landscape. Given that static identification takes no account of the movement of people through a city, this information alone is inadequate to estimate the ecosystem services that urban dwellers receive and new sources of data are required to progress from static to dynamic relationships between urban people and spaces.

This chapter demonstrates that crowdsourced data from Twitter and BetterPoints can be successfully employed to investigate dynamic nature of human movement through an urban landscape and draws on numerous examples and applications of the utility of this information to urban planners. Both Twitter and BetterPoints data were successfully used to identify spatial patterning and movement of individuals through the urban landscape of Birmingham, UK. While the static nature of geolocated social media posts, such as those derived from Twitter, limits their utility in applications beyond identifying hotspots of activity in urban green spaces, data from physical activity recording apps, such as BetterPoints, have utility for a number of applications. These include the identification of exact route ways taken through urban green spaces, examining the mobility of certain subgroups such as commuters, improving the efficiency of urban mobility schemes and assessing the accessibility of urban green spaces and identifying access inhibitors. Data from both apps and social media may also have utility in informing theoretical debates surrounding conceptualisations of the relationships between people and space.

A number of features of social networks and app data make them well placed to be used as sources of data in urban research. These include the ubiquity of smart device and app use among urban citizens leading to vast numbers of data points being generated, the high spatial resolution of these data points and the non-obtrusive method of participant observation which the data-sharing aspect of these devices and apps enable.

Despite this, a number of limitations remain with using geolocated data generated in this way to explore the movement of individuals through urban landscapes. These include issues with user privacy, the lack of user demographic metadata and the substantial computational power required to interpret and analyse the vast outputs of social media and app derived data.

However, these limitations to not detract from the fact that app and social media derived data show much potential in providing information to researchers and urban planners about the movement of individuals through urban landscapes; which is turn is an important first step to understanding how the dynamic relationship between urban people and spaces impacts upon ecosystem service receipt.

Chapter 8. Synthesis and General discussion

This thesis has sought to explore the utility of new data sources and methods in understanding how people use and consume urban green spaces, and thus receive the cultural ecosystem services these spaces provide. Two explicit aims of this thesis were presented at the outset:

(i) to demonstrate the utility of data obtained from smart device enabled platforms (social networks and apps) in understanding socio-ecological interactions in urban areas between human populations and urban green spaces;
(ii) to critically evaluate the use of these data sources for researchers and policy makers and to evaluate how useful the data provided is to them.

This chapter begins with a synthesis of the information presented in this thesis, drawing together the different types of investigations presented using data obtained from Twitter. Through outlining the results and conclusions of the previous chapters, it presents a summary of the information that this thesis has demonstrated can be obtained from social media and apps to inform investigations into urban socio-ecological interactions. The utility of this information for urban planners is then outlined, and the advantages and limitations of the use of this type of data are discussed.

8.1 Synthesis

As urban environments continue to expand on a global scale (Seto et al., 2011), they are inhabited and experienced by an increasing number of human individuals (Wu, 2008; UNHABITAT, 2016). There is much discussion in a range of research disciplines, surrounding how to improve the liveability and sustainability of cities, and how to enhance the quality of life for their populations. To do this, the processes ongoing within the urban system, and how human individuals interact with and experience the urban landscape around them must first be understood (Daniel et al., 2012). Such an effort requires people derived, city specific and substantive dataset on the interactions occurring within the cityscape (Kallus, 2010). While technology has been embraced in many realms of urban life, leading to the concept of the ‘smart city’ gaining rapid traction in academia (Chourabi et al., 2012; Kitchin, 2014; Albino et al., 2015; Gaiani et al., 2017), business (Söderström et al., 2014) and government (Neirotti et al., 2014; Meijer and Bolívar, 2016); ecosystem service research has yet to align itself with this new way of conceptualising urban spaces and to realise the benefits to data collection that connected, digital urban spaces can provide to researchers. This thesis pioneered the use of crowdsourced data derived from smart device applications in investigations into soico-ecological interactions occurring within urban green spaces. In doing so, it has demonstrated that these new sources of data offer vast opportunities for the real-time analysis of city life; and particularly for cultural ecosystem service research, which has typically lacked exploration given the difficulty in generating appropriate datasets for study.

This thesis evaluated how Twitter data can be used to investigate interactions between people and urban green spaces, with each chapter addressing a specific objective. Chapter Three demonstrated how Twitter data confirm the notion that urban green spaces are an important resource for human populations, and provide a range of cultural ecosystem services to urban populations (Groenewegen et al., 2006; Westphal, 2003; Seeland et al., 2009, Kim and Kaplan, 2004). Indeed, this chapter found Twitter data provided a viable approach to investigate human interactions with urban green spaces. The Twitter centred methodology presented identified urban green spaces as important locations for organised events, presenting opportunities for social interaction, economic opportunity and fostering community identity. Thus, consolidating the notion that spaces can have multiple, simultaneous functions for different groups of users (Peng et al., 2016). The utilisation of this technique has the potential to be more cost and time efficient than previous methodologies, as well as enabling longitudinal study through space and time at greater spatial and temporal resolutions.

Given increasing levels of obesity, urban green spaces are important locations providing space and opportunity for individuals to exercise and reduce their risk of obesity related disease (Han et al., 2103; Bedimo-Rung et al., 2005). To understand how the benefits of outdoor physical activity in urban green spaces are transferred to human populations, consideration must be given to when people are using them, what they are using them for and what factors may affect the use of space. Chapter Four examined the extent to which Twitter data can provide this information, in an assessment of reported physical activity engagement in urban green spaces. In doing so, this chapter evaluated a specific use of Twitter data in investigating the temporality of this cultural ecosystem service. Using Twitter data, variance was identified in reported physical activity engagement between two seasons. A number of factors, including meteorology, park characteristics and amenities and the role of organised sports events were explored in order to explain these findings. Understanding how physical activity engagement in urban green space varies seasonally is important in ensuring policy interventions to increase physical activity are targeted most effectively (Beighle et al., 2008).

Chapter Five investigated the potential of using Twitter data as a source of sentiment and emotion responses information. Interactions between humans and nature are understood to be beneficial for human mental and emotional well-being (Wilson, 1984; Chiesura, 2004; Fuller and Gaston, 2009). In cities, urban green spaces provide urban populations with key opportunities for contact with nature (Kremer et al., 2016). This chapter investigated the emotions experienced by individuals whilst using urban green spaces. Both positive and negative responses were identified from the tweet texts and their seasonal variation was investigated. Positive responses were found to be more common than negative responses across all seasons, with happiness and appreciation of surroundings being the most common positive emotions identified. For the negative responses, fear and anger were present in similar amounts with smaller annotations of sadness and disgust also present. The results presented in this chapter support existing research, and suggest that crowdsourced data offer a new way to progress methodological approaches in understanding human interaction with urban green space. Tweets were successfully analysed to identify sense of place, attachment, aesthetic and other factors pertaining to the experience of human-environment interactions in urban green spaces.

Extracting the sentiment present in tweets is increasingly recognised as a valuable approach to gathering information on the mood, opinion and emotional responses of individuals in a variety of contexts (Zhang et al., 2011), and urban research is no exception (Bertrand et al., 2013; Martinez and Gonzalez, 2013). However, there remains no standardised approach to extract the sentiment form tweet text with numerous methodologies are employed to this end. Chapter Six presented a first comparison of three different methods of sentiment analysis in an urban research context. When used on the same corpus of tweets, manual, semi-automated and automated methods were found to generate similar numbers of positive/negative/neutral annotated tweets. However, inter-method consistency in tweet assignment between the methods was low. A number of limitations were found to be associated with each method and caution is advised in the adoption of each method in future research.

Rather than an empirical application of data, Chapter Seven presented a discussion exploring the potential utility of crowdsourced data from Twitter and fitness based apps in investigating spatial and temporal fluxes in human movement through cities. Ecosystem service research has typically relied on static identification of the localities of provisioning based on ecosystem structure and function. However, this information alone is inadequate to estimate the ecosystem services that urban dwellers receive as it takes no account of the movement of individuals through the urban landscape. Such movement is critical as the routes chosen by individuals will impact on the interactions they have with urban green spaces, and the ecosystem services they may or may not receive as a result. The discussion presented identifies numerous opportunities presented by crowdsourced data from social networks and apps in furthering the ability of researcher to investigate the dynamic relationship between urban individuals and the urban green spaces they pass through, or near as they navigate the cityscape; with these investigations having practical application in urban planning and decision making.

The applications of Twitter and, to a lesser extent, app data presented in these chapters address the first aim of this thesis: to demonstrate the utility of data obtained from a smart device enabled platforms in understanding a variety of socio-ecological interactions in urban areas between human populations and urban green spaces. In doing so, this thesis proved Twitter data can provide a range of information about urban citizens and how they interact with urban green spaces, which can be used as a first step to understanding cultural ecosystem service provision.

8.2 Utility of crowdsourced data to urban planners

This thesis identified that a range of information can be obtained from Twitter and other app data, with this information having significant utility to urban planners and park managers.

The green space action cycle (WHO, 2017) is a useful depiction of the processes required to provide functional urban green spaces which cater for the range of local user needs (Figure 8.1).

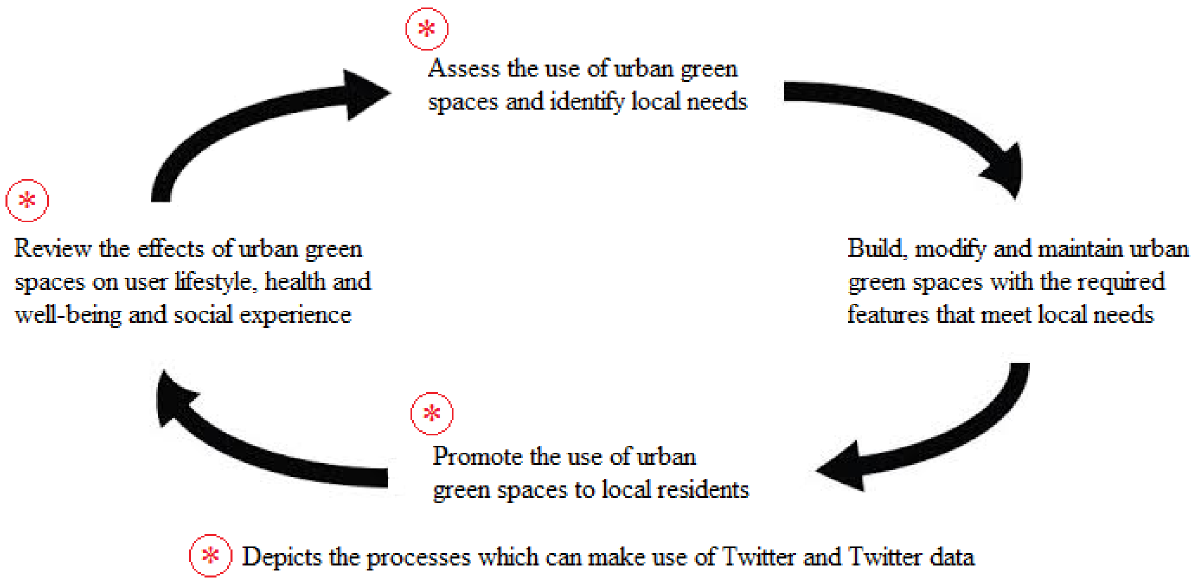


Figure 8.1 *The green space action cycle adapted from WHO (2017). Red asterisks indicate the stages of the cycle where Twitter can be used to promote, monitor or gather information to aid the process.*

This thesis has demonstrated that Twitter can be used at three stages of the green space action cycle to provide information gathering, monitoring and promotion tools to urban planners and park managers to ensure the delivery and maintenance of urban green spaces which meet the needs of local urban communities. Twitter and other apps and other social networks can be used to assess the use of urban green space by urban individuals. This information can be used to inform the building, modification of urban green space to ensure they meet the needs of their users. Twitter data can also be used to review the effects of interactions with urban green spaces on user lifestyle, health, well-being and social experience.

As well as providing new sources of data, social networks provide opportunities for decision makers to connect and communicate with citizens. Social networks and apps have a significant role to play in supporting urban planning and the promotion of green spaces and nature in cities. Promoting activities to local residents over these platforms will help increase the use and awareness of the variety of opportunities in urban green space. Social networks are also important platforms which can facilitate increased communication between governments and citizens, enable knowledge gathering and sharing between stakeholders and increasing public participation within urban planning, governance and decision making (Guerrero et al., 2016). The ability to create and access tweets containing specific hashtags also presents opportunities for targeted data collection.

The ability to spatially position information obtained from Twitter and apps through the geolocation feature inherent to smart devices, provides the opportunity to map cultural ecosystem services. This could give a more complete picture of ecosystem service provision in urban areas and begin to redress the balance of the predominance of regulating and provisioning services in ecosystem service research (Daniel et al., 2012). Integrating cultural ecosystem services into urban planning has been previously problematic due to their intangible and subjective nature and the difficulty in attributing value to them (Milcu et al., 2013). However, this thesis illustrates that crowdsourced data from social networks and apps can provide information about urban socio-ecological interactions (Table 8.1) and thus the cultural ecosystem services people are experiencing. City authorities could use these data as a driver for protection and investment of urban green spaces as the information comes directly from those benefiting from the cultural ecosystem services they provide.

Table 8.1 *A summary of the information gathered from social network and app data in this thesis and its potential use for urban planners and park managers.*

Chapter	**Type of information gathered**	**Potential use**
Three	The diversity of organised events occurring in urban green spaces.	• Fuller understanding of the importance of urban green spaces to local communities • Awareness of the variety of users can aid in the design of user focused, functional urban green spaces (Buchel and Frantzeskaki, 2015) • Provision of justification for the continued presence of urban green space in light of pressure for development • Promotion of cultural, social and business opportunities and events taking place in urban green spaces
Four	Seasonal and temporal variation in reported outdoor physical activity engagement, for both independent and organised physical activities.	• Monitoring the success of organised sports events • Provision of justification the continued provision of organised sports events when funding may be limited and an evidence based is required • Identification of where and when reported physical activity engagement is reduced to enable targeted strategies to help this • The identification of successful organised sports could enable similar events to be replicated in other locations • Promotion of organised physical activity events taking place in urban green spaces
Five	Emotional responses of individuals to the urban green spaces they experience.	• Identification of the causes of certain experienced emotions • Enables the identification of the features/events which cause positive response so they can be maintained or replicated

		• Enables the identification of the features/events which cause negative responses so they can be improved or removed
Seven	Movement of urban individuals through urban landscapes	• Understanding the dynamic relationship between people and urban green spaces • Enables more accurate spatial and temporal identification of the location of cultural ecosystem service provision and receipt • Quantification of cultural ecosystem service provision and receipt, shifting the balance in ecosystem service research from predominantly regulating and provisioning services to cultural services • Assessment of the efficiency of urban mobility schemes • Identification of popular route ways, providing an evidence base for implementing improvements such as cycle lanes, green infrastructure or traffic calming measures

In line with the second aim of this thesis, an evaluation of using data from social networks and apps has been provided in each application, to provide holistic and contextualised findings and conclusions. From these, a summary of the benefits and limitations associated with the use of Twitter data in an urban research context have been collated.

8.3 The importance of Twitter data in an urban context

Identified throughout this thesis, Twitter provides a source of data with no associated financial cost. Indeed, the datasets and methods of obtaining them used in this thesis were all free, demonstrating that crowdsourced data from this social network has the potential to gather large amounts of data in an inexpensive manner (Illingworth et al., 2014). Where budgets may be constrained this is a significant advantage in using this type of data for both researchers and policy makers.

In a similar vein, the data proved to be easily accessible via the Twitter API. A simple code affords access a corpus of tweets, and a number of programmes already have the necessary coding functions in place. The ease of accessing a corpus of tweets provides an easily replicated methodology, improving the potential for a standardised approach of data collection to be developed.

A number of attributes associated with Twitter make it a convenient source of data for a range of analyses. The consistency of Twitter data, in that tweets are limited to 140 characters, makes analysis more straightforward compared to other social media posts of varying length (Highfield and Lever, 2014). The metadata downloaded alongside tweet text also enables the use of Twitter data in a number of research applications (Demirbas et al., 2010; Pettit et al., 2012). The geolocation, time and date provided with each tweet present opportunities for exploring spatial and temporal patterning, making it an ideal source for geographic investigation. Indeed, the geolocation feature enables users to bind together embodiments and mobility with place and memory (Kalin and Frith, 2016) resulting in an affective response to space that can be captured and explored by researchers.

Twitter is a popular social network and has established a vast user base, making it possible to study the information made available by a large number of individuals. A huge volume of data points are continually created by its users, making it a suitable source of data for quantitative investigations and analyses. Indeed, Chapter Four of this thesis demonstrated the potential of Twitter data in statistical analysis where it was used conjunctively with data from other sources (Netatmo) to inform investigations. As well as quantitative based investigations, Chapter Five of this thesis validated Twitter as a source of data for qualitative analysis. The information contained in tweet text was successfully used as the basis for sentiment analysis. This thesis posits that when used correctly and in the appropriate contexts, Twitter data provide the opportunity for both quantitative and qualitative analyses in the urban research context.

Using Twitter data also provides an unobtrusive method of non-participation and observation. As shown in this thesis, it can be used successfully for constant monitoring, enabling the generation of continual datasets over time. This is a benefit of Twitter data when compared to more traditional methods of investigating urban socio-ecological interactions, for example surveys and questionnaires, which tend to be carried out over short periods of time due to time and financial constraints. Cultural ecosystem services such as those investigated in this thesis often vary with season and to investigate delivery and receipt most accurately, high frequency observations are needed to identify this temporal variation (Buchel and Franzteskaki, 2015). The built environment is a complex system which experimental and observational approaches may fail to capture. Urban green space is notoriously heterogeneous differing in size, quality and range of facilities (Wolch et al., 2014) meaning blanket solutions to improving access and use may not be appropriate, and consideration should be given to local contexts. Thus, methods such as the one proposed in this thesis, which can successfully study high spatial and temporal variation in the use of urban green spaces are needed to identify where interventions may be needed. Public Health England has explicitly stated the need for methodological development in research to better investigate the processes ongoing within an urban system (PHE, 2017). The use of novel sources of data such as Twitter and BetterPoints provide an attempt at such progression.

Twitter data can also be used to provide information at a high spatial resolution and thus a localised level. Within urban planning, this is particularly useful given that cities and their populations are highly heterogeneous. High resolution data are required to fully capture this variation and enable the development of planning interventions that are contextualised and appropriate for the location in which they are being implemented. This is particularly relevant to urban green spaces as these are often used as locations through which specific local circumstances are addressed (Handley et al., 2003), making localised data on such spaces important. For example, as identified in Chapter Three, urban green spaces can be used to tackle social exclusion and isolation by enabling and promoting greater use of green space by certain groups including refugees, women and the disabled. High resolution data could be used to monitor the success of various schemes aimed at improving specific circumstances.

In relation to the study of emotional responses of individuals to urban spaces, Aiello et al., (2016) recognise that there has been no way to capture the emotions of individuals at scale, making them hard to incorporate into city planning. Georeferenced or place labelled Twitter data can help to address the challenge. Chapter Five illustrated that emotion can be successfully identified in such datasets, and there is potential to upscale these datasets and utilise the emotional information they contain.

Finally, the findings presented in the chapters of this thesis corroborate with the findings of previous research in each case. This alone goes some way to validating the use of Twitter data in urban research and explorations into cultural ecosystem service provision by urban green spaces. Indeed, Twitter data could be used to complement traditional methods of data collection, to create more robust, spatially referenced datasets. Green spaces within communities are critical spaces in the urban landscape for facilitating socialisation and interactions beyond an individual’s own social circle, through recreational and cultural activities such as picnics and festivals (Smith et al., 2013); a notion consolidated in the thematic analysis of Chapter Three. Similarly, the significant role of urban green spaces in facilitating engagement with outdoor physical activity and active recreation (Bedimo-Ring et al., 2005; Maas et al., 2008; McCormack et al., 2010) was identified in Chapter Four. In line with the findings of previous research, the outcomes of Chapter Five found urban green spaces to elicit significant positive responses from individuals, confirming their beneficial role in improving the emotional and psychological well-being of urban individual (Barton and Pretty, 2010; Shanahan et al., 2016).

8.4 Limitations of Twitter data in an urban context

Despite the benefits identified with the use of Twitter data in an urban research context, this thesis also identified a number of limitations of Twitter as a source of data. In order for Twitter data to be utilised most effectively in future research endeavours, it is important for these limitations to be discussed openly; enabling methodologies to be developed which account for and overcome these limitations.

While this thesis has illustrated that Twitter could provide large enough datasets to enable robust analysis, it is important to realise the limitations inherent in these datasets, due to the nature of downloading tweets via the Twitter REST API. Tweets made available for download by Twitter through their REST API are a randomised 1% sample of all the tweets related to the specific search query. Therefore, any dataset collated in this way is by no means complete.

Methods such as those presented in this thesis rely intrinsically on crowdsourced data generated from smart devices. Given the reliance of such methods on infrastructures of internet connectivity, limitations on the data available for capture should be considered.

Internet connectivity can vary substantially between network providers with signal being intermittent, although this is rare in urban areas. Areas with limited or no internet connectivity may lead areas with no recorded use, which may not accurately reflect reality (Chatzimiliodis et al., 2012). The research presented in this thesis takes place in the metropolitan area of Birmingham which has widespread 4G coverage. For this reason infrastructure restrictions are unlikely to be relevant to this research. However, this limitation should be considered if employed in other areas as limited internet connectivity may restrict the size of the dataset available to researcher.

Much research undertaken to explore urban social processes requires knowledge of the population demography included in the sample group. This information enables nuanced appreciation of the influential cultural dynamics which are operational in forming green space experience. However, Twitter does not provide demographic information about its users; and while this has not inhibited the investigations presented in this thesis it is important to acknowledge. A lack of demographic information about the users in a sample population may restrict the utility of Twitter data for certain types of research, particularly cultural and social investigations.

Using crowdsourced information from Twitter immediately limits the population one can investigate. Research identifies that the Twitter population in a highly non-uniform sample of the population (Mislove et al., 2011), with the elderly showing disproportionate levels of non-engagement with this technology (Zickuhr and Madden, 2012) and therefore excluded from a dataset obtained from Twitter. Members of the urban population who do not own a smart device or are not a user of Twitter are excluded from the sample population. Users of social media therefore, do not reflect the diversity of the urban populations they may be employed to investigate. Given that urban green spaces are promoted as locations where all members of the community can come together, Twitter data may not be the most appropriate capture the full plethora of socio-ecological interactions if used in isolation from other data sources.

The biases inherent in using a dataset derived from the users of a social network such as Twitter, will inevitably have an effect on the outcomes of the research. This thesis has identified the generic limitations associated with these user base biases however; more direct consideration needs to be given as to how they may affect the results of the applications presented herein. For a study attempting to identify the complete range of events ongoing within a park (Chapter three) it may be that the lack of certain demographic groups from the Twitter user base, for example older people, may result in a lack of events targeted as this group. If there is little engagement of elderly people with Twitter, it stands to reason that these events would not be advertised, mentioned or discussed on Twitter. Therefore certain events may be missed from analysis leading to incompleteness in the narrative of green spaces as functional spaces for social interaction. For studies using Twitter data to provide information on the range of physical activities engaged with in a park (Chapter 4), the biases in the Twitter user base may lead to overrepresentation and underrepresentation of certain activities. For example, if one activity is popular with a specific subgroup that is very active on Twitter, then this will be overrepresented in the resultant data compared to a different activity which may be just as popular with another subgroup, but who are less active on Twitter. Similarly, for studies attempting to use Twitter data as the input for sentiment analysis and other emotional explorations (Chapters five and six) the biases inherent in the Twitter user base will effect. In cases where this emotional information is used in isolation to inform decisions as to the features that illicit positive and negative responses, one would end up designing parks to cater for the users of Twitter, rather than the general population. As is highlighted throughout this thesis, the Twitter user base does accurately reflect the general population so if planners are to use this information to inform decision making this fact should not be forgotten and measures should be put in place to ensure this information adds to, rather than replaces information obtained from more representative samples of the population.

There remains limited investigation into the sub-groups of users of green space with most instead focused on populations as a whole. For example children, the elderly or homeless people may interact and receive ecosystem services/disservices differently, but as of yet this has not been thoroughly investigated. This thesis suggests therefore that Twitter data are not appropriate for research intending to investigate variations in the use of urban spaces by various sectors of the population. However, application derived data such as that used in Chapter Seven could fill this data gap if a user is asked to specify a number of personal details. For example, the BetterPoints application requires their users to submit age and gender information as standard.

This thesis identifies Twitter data as a viable source of data for qualitative analysis; however, the lack of any standardised approach currently restricts the potential utility of Twitter data in urban research. As identified in Chapter Six, there remain numerous ways to analyse the sentiment present in tweet text. A comparison of three methods of sentiment analysis; manual, automated and semi-automated, was undertaken on a corpus of tweets. Despite the three methods annotating similar numbers of tweets as positive/neutral/negative, inter-method consistency in tweet assignment between the methods was found to be low. This chapter consolidates the requirement for caution when interpreting the outputs of qualitative analysis undertaken on tweet text and suggests that the use of technology to aid annotation should be treated with caution.

For information about socio-ecological interactions in urban green spaces to be useful for urban planners and policy makers, it is important that it is contextualised. In short, this requires an understanding of how individuals respond and behave in the wider urban environment in order to identify if these differ in urban green space compared to the rest of the urban landscape. The assessment of sentiment present in tweet text relating to urban green spaces presented in Chapter Five of this thesis evidences that Twitter can be successfully used as a source of data to inform on the emotional responses of people while they experience and interact with urban green spaces. Given that this thesis presents a first attempt at using Twitter to explore these sentiments, there is currently no control study to which the outputs of this thesis can be compared. Thus, while Chapter Five demonstrated the range of emotional information and their causes that can be captured from tweets, further research is needed to explore how the sentiments of other urban locations are portrayed on by users of Twitter.

Finally, consideration should be given to the ethical discussion surrounding the use of data generated by the users of social media. While much of the data is generated by users and placed by them into the public domain, there may often be a lack of awareness of how the information they create and make public can be used by a variety of individuals and organisations. Indeed, just because the data is made available, doesn't necessarily mean that it is ethical or morally appropriate to do so (Boyd and Crawford, 2012). Numerous ethical questions are raised by the use of social media data; for example is it right to analyse the information of an individual if informed consent has not been given? Whilst it may be unreasonable to expect researchers to obtain consent from all the users of social media whose information they utilise, it may become increasingly problematic for researcher to justify the use of such data as ethical simply because the data is publically accessible.

Any data on human subjects inevitably raises privacy issues and ethical concerns, particularly when those individuals do not know what the information they generate will be used for. It is likely that none of the users who generated the information used in this thesis are aware that their information has been used in this way. And while the applications demonstrated in this thesis are benign and look to use the information for good and to improve their experience of urban green spaces, other research endeavours may be less well-meaning.

Given the ubiquity of social media use among vast swaths of the population, it is concerning that there is a lack of awareness among the average user as to how the information they generate may be used. An increase in educational effort to inform users of how the information they generate can be used and the variety of individuals and organisations that can access their public data may perhaps be a useful starting point to enable the use of this data to be more ethical. Only once an individual is fully aware of how their information can be used, does it seem fully appropriate to use it for research purposes, particularly those which may be more morally dubious than those presented in this thesis.

Chapter 9. Conclusion

Urban areas are defined by their human presence, and the urban landscape provides an important backdrop against which the human and natural components of the system interact. As such, ecosystem services are co-produced and a socio-ecological approach to understanding urban ecosystem services is critical to furthering governance, planning and policies that seek to utilise ecosystem services for a variety of goals including public health and urban sustainability. Cultural ecosystem services are a perfectly placed to bridging the gap between different disciplines and research agendas. Given their relatedness to human wellbeing, attitudes and beliefs, cultural ecosystem services highlight powerful linkages between the social and ecological sciences, providing a framework that researchers from numerous disciplines can engage with to gain a holistic understanding of the socio-ecological interactions resulting from human contact with urban ecosystems. This thesis provides strong evidence for the utility of an interdisciplinary approach for urban research and provides an important contribution by identifying the role of crowdsourced social media and app data in investigating urban socio-ecological interactions. It has demonstrated the broad range of information that can be obtained from these data sources to inform a variety of quantitative and qualitative investigations into the cultural ecosystem services received by urban citizens from urban green spaces.

This thesis has presented a series of applications using crowdsourced data to investigate a range of socio-ecological interactions occurring within urban green spaces. Numerous cultural ecosystem services are facilitated by the presence of green spaces in the urban landscape, which this thesis has demonstrated can be successfully investigated and monitored using crowdsourced data from social media and apps. This thesis confirms the use of such data presents methodological progression in the capture of urban socio-ecological interactions

A number of benefits associated with the use of crowdsourced social media and app data have been identified. These include the high spatial and temporal resolution of the data, making them highly valuable in explorations of the spatial and temporal variation in urban green space use and ecosystem service provision and receipt. Additionally, the potential sizes of the datasets which are freely available present opportunities to gather large amounts of data in an inexpensive manner. The versatility of these data in both quantitative and qualitative terms has also been demonstrated by this thesis, presenting opportunities for its incorporation in a plethora of urban research endeavours.

However, there remain limitations to the use of crowdsourced data, particularly in the context of geographical research, given the lack of user demographic metadata available to researchers. This significantly inhibits its utility in understanding the influential cultural dynamics which are operational in forming green space experience. Inherent age, gender and class biases in the users of social networks and apps are also an important consideration, and may limit the applicability of findings to wider populations. Additionally, the reliance of social media and apps on technical infrastructures influences the generation of data and may limit its utility in specific locations.

Throughout this thesis, the relevance and utility of these data to urban planners has been discussed. The information made available by users of social media and other apps enables the provision of evidence based assessments into the role of urban green spaces. This can be used to justify their continued presence in the urban landscape despite substantial pressure from development. Furthermore, Twitter textual information can be used to identify issues or problem areas in green spaces which need addressing; and enables the identification of the features which promote positive responses and increase the likelihood of green space use. Once identified, these can be maintained in that location and replicated in other spaces to promote use in other urban green spaces.

Alongside the role of providing data to inform urban planning decision making, this thesis identified that social media and apps present an important interface which can facilitate communication between the city and its citizens. Their use to this end can enable knowledge gathering and sharing between stakeholders, and can be used to increase public participation within urban planning, governance and decision making. Through the provision of data, and as a platform to facilitate increased communication between stakeholders, this thesis has identified that crowdsourced data from social networks and apps have a significant role in supporting the green space action cycle and thus, have a significant role in enabling the provision of functional urban green spaces which cater for the range of local user needs. Through the identification of the implications and limitations of the use of these data sources, a foundation for future work using crowdsourced data from social networks and apps in an urban context has been established.

10. List of References

Aarts, H., Paulussen, T. and Schaalma, H. (1997) Physical exercise habit: on the conceptualization and formation of habitual health behaviours *Health Education Research* **12**: 363-74.

Agryzkov, T., Marti, P., Nolasco-Cirugeda, A., Serrano-Estrada, L., Tortosa, L. and Vicent, J. (2016) Analysing successful public spaces in an urban street network using data from social networks Foursquare and Twitter *Applied Network Science* **1**: 12.

Aiello, L., Schifanella, R., Quercia, D. and Aletta, F. (2016) Chatty maps: constructing sound maps of urban areas from social media data *Open Science* **3 (3)**: 150690.

Albino, V., Berardi, U. and Dangelico, R. (2015) Smart cities: Definitions, dimensions, performance, and initiatives *Journal of Urban Technology* **22 (1)**: 3-21.

Anderson and Smith (2001) Editorial: Emotional Geographies *Transactions of the Institute of British Geographers* **26 (11)**: 7-10.

Annesi, J. (2002) Self-motivation moderates effect of exercise-induced feelings of adherence *Perceptual and Motor Skills* **94 (2)**: 467-475.

Ash, J., Kitchin, R. and Leszczynski, A. (2016) Digital Gegraphies, digital geogrpahies? *Progress in Human Geography*: 0309132516664800.

Bagstad, K., Johnson, G., Voigt, B. and Villa F. (2013) Spatial dynamics of ecosystem service flows: A comprehensive approach to quantifying actual services *Ecosystem Services* **4**: 117-125.

Barbosa, L. and Feng, J. (2010) Robust sentiment detection on twitter from biased and noisy data, In: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. 36-44. Association for Computational Linguistics.

Barbosa, O., Tratalos, J., Armsworth, P., Davies, R., Fuller, R., Johnson, P. and Gaston, K. (2007) Who benefits from access to green space? A case study from Sheffield, UK *Landscape and Urban Planning* **83 (2)**: 187-195.

Barton, J. and Pretty, J. (2010) What is the best dose of nature and green exercise for improving mental health? A multi-study analysis *Environmental science & technology* **44 (10)**: 3947-3955.

Bastian, O., Haase, D. and Grunewald, K. (2012) Ecosystem properties, potentials and services–the EPPS conceptual framework and an urban application example *Ecological Indicators* **21**: 7-16.

Bedimo-Rung, A., Mowen, A. and Cohen, D. (2005) The significance of parks to physical activity and public health: A conceptual model *American Journal of Preventative Medicine* **28 (2)**: 159-168.

Beighle, A., Alderman, B., Morgan, C. and Masurier, G. (2008) Seasonality in children's pedometer-measured physical activity levels *Research Quarterly for Exercise and Sport* **79 (2)**: 256-260.

Belza, B., Walwick, J., Shiu-Thornton, S., Schwartz, S., Taylor, M. and LoGerfo, J. (2004) Older adult perspectives on physical activity and exercise: voices from multiple cultures *Preventing Chronic Disease* **1 (4)**: A09.

Bendler, J., Wagner, S., Brandt, T. and Neumann, D. (2014) Taming uncertainty in big data evidence from social media in urban areas *Business and Information Systems Engineering* **6 (5)**: 279-288.

Bennett, E., Peterson, G. and Gordon, L. (2009) Understanding relationships among multiple ecosystem services *Ecology Letters* **12**: 1394-1404.

Berman, M., Jonides, J., and Kaplan, S. (2008) The cognitive benefits of interacting with nature *Psychological science* **19 (12)**: 1207-1212.

Bertrand, K., Bialik, M., Virdee, K., Gros, A. and Bar-Yam, Y. (2013) Sentiment in New York city: A high resolution spatial and temporal view *arXiv preprint arXiv:1308.5010*.

Bifet, A. and Frank, E. (2010) Sentiment knowledge discovery in twitter streaming data, In: *DS proceedings of the 13th international conference on discovery science*, 1-15 Springer-Verlag Berlin, Heidelberg.

Birmingham City Council (BCC) (2006) *The Future of Birmingham's Parks and Open Spaces: Supplementary Planning Document*. BCC, United Kingdom.

Birmingham City Council (BCC) (2011) *2011 Population Census: Quick Reference Set of Key Variables, Birmingham City Council: Population and Growth Strategy*. Birmingham City Council, United Kingdom.

Birmingham City Council (BCC) (2013) *Green Living Spaces Plan*. BCC, United Kingdom.

Birmingham City Council (BCC) (2014) *Tackling childhood obesity in Birmingham.* Report of the Health and Social Care Overview and Scrutiny Committee, United Kingdom.

Birmingham City Council (BCC) (2016a) Parks and nature conservation http://www.birmingham.gov.uk/parks Accessed 04/05/2016

Birmingham City Council (BCC) (2016b) *Birmingham City Council Service Reviews, Green Paper: Sport and Physical Activity Final Version*. Birmingham City Council, Birmingham.

Bixler, R. and Floyd, M. (1997) Nature is scary, disgusting, and uncomfortable *Environment and behavior* **29 (4)**: 443-467.

Bixler, R., Carlisle, C., Hammltt, W. and Floyd, M. (1994) Observed fears and discomforts among urban students on field trips to wildland areas *The Journal of Environmental Education* **26 (1)**: 24-33.

Bjork, J., Albin, M., Grahn, P., Jacobsson, H., Ardo, J., Wadbro, J., Ostergren, P. and Skarback, E. (2008) Recreational values of the natural environment in relation to neighbourhood satisfaction, physical activity, obesity and wellbeing *Journal of Epidemiology and Community Health* **62 (4):** e2.

Bjornsdottir, G., Arnadottir, S. and Halldorsdottir, S. (2012) Facilitators of and barriers to physical activity in retirement communities: experiences of older women in urban areas *Physical Therapy* **92 (4)**: 551-562.

Boleyn, T. and Honari, M. (2005) *Health Ecology: Health, culture and human-environment interaction*. Routledge, London.

Bollen, J., Mao, H. and Zeng, X. (2011) Twitter mood predicts the stock market, *Journal of Computational Science* **2 (1)**: 1–8.

Bolund, P. and Hunhammar, S. (1999) Ecosystem services in urban areas *Ecological economics* **29 (2)**: 293-301.

Bookchin, M. (1988) Social ecology versus deep ecology *Socialist Review* **18** (3): 9-29.

Botkin, D. and Beveridge, C. (1997) Cities as environments *Urban Ecosystems* **1**: 3-19.

Boy, J. and Uitermark, J. (2016) How to study the city on Instagram *PloSone* **11(6)**: e0158161.

Boyd, D. and Crawford, K. (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon *Information, communication & society* **15 (5):** 662-679.

Brand, J. (2012) *The crowdsourcing handbook – Everything you need to know about crowdsourcing.* Emereo Publishing, Brisbane.

Bratman, G., Hamilton, J. and Daily, G. (2012) The impacts of nature experience on human cognitive function and mental health *Annals of the New York Academy of Sciences* **1249 (1)**: 118-136.

Brereton, F., Clinch, J. and Ferreira, S. (2008) Happiness, geography and the environment *Ecological Economics* **65 (2)**: 386-396.

Broach, J., Dill, J. and Gliebe, J. (2012) Where do cyclists ride? A route choice model developed with revealed preference GPS data *Transportation Research Part A: Policy and Practice* **46 (10)**: 1730-1740.

Brockman, R., Jago, R. and Fox, K. (2011) Children's active play: self-reported motivators, barriers and facilitators *BMC Public Health* **11 (1)**: 461.

Brodersen, N., Steptoe, A., Williamson, S. and Wardle, J. (2005) Sociodemographic, developmental, environmental, and psychological correlates of physical activity and sedentary behavior at age 11 to 12 *Annals of Behavioral Medicine* **29 (1)**: 2-11.

Brownson, R., Baker, E., Housemann, R., Brennan, L. and Bacak, S. (2001) Environmental and policy determinants of physical activity in the United States *American Journal of Public Health* **91 (12)**: 1995-2003.

Bruns, A. and Burgess, J. (2011) The use of Twitter hashtags in the formation of ad hoc publics, In: *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011.*

Bryman, A. (2001) *Social Research Methods*. Oxford University Press, New York.

Bubolz, M. and Sontag, M. (1993) Human ecology theory, In: *Sourcebook of family theories and methods*, 419-450. Springer, United States.

Buchel, S. and Frantzeskaki, N. (2015) Citizens' voice: A case study about perceived ecosystem services by urban park users in Rotterdam, Netherlands *Ecosystem Services* **12**: 169-177.

Burghardt, T., Buchmann, E., Müller, J. and Böhm, K. (2009) Understanding user preferences and awareness: privacy mechanisms in location-based services, In: *On the Move to Meaningful Internet Systems: OTM 2009*, 304-321. Springer, Berlin Heidelberg.

Burton, E. (2002) Measuring urban compactness in UK towns and cities *Environment and Planning B: Planning and Design* **29**: 219-250.

Cadenasso, M., Pickett, S. and Schwarz, K. (2007) Spatial heterogeneity in urban ecosystems: conceptualising land cover and a framework for classification *Frontiers in Ecology and Environment* **5**: 80–88.

Carpenter, S., Mooney H., Agard, J., Capistrano, D., DeFries, R., Diaz, S., Dietz, T., Duraiappah, A., Oteng-Yeboah, A., Pereira, H., Perrings, C., Reid, W., Sarukhan, J., Scholes, R. and Whyte, A. (2009) Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment *PNAS* **106 (5)**: 1305-1312.

Catton, Jr. and Dunlap, R. (1978) Environmental sociology: a new paradigm *The American Sociologist* **13** (1): 41-49.

Chaix, B., Meline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C. and Kestens, Y. (2013) GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? *Health & Place* **21**: 46-51.

Chan, C., Ryan, D. and Tudor-Locke, C. (2006) Relationship between objective measures of physical activity and weather: a longitudinal study *International Journal of Behavioural Nutrition and Physical Activity* **3 (21)**: 21.

Chapman, L., Bell, C. and Bell, S. (2017) Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations *International Journal of Climatology* **37 (9)**: 3597-3605.

Chapman, L., Muller, C., Young, D., Warren, E., Grimmond, C., Cai, X-M., Ferranti, E. (2015) The Birmingham Urban Climate Laboratory: An open meteorological testbed and challenges of the smart city *Bulletin of the American Meteorological Society* **96**: 1545-1560.

Chatzimiliodis, G., Konstantinidis, A., Laoudias, C. and Zeinalipour-Yatzi, D. (2012) Crowdsourcing with smartphones *Internet Computing – IEEE* **16 (5)**: 36-44.

Cheong, M. and Lee, V. (2011) A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter *Information systems frontiers* **13 (1)**: 45-59.

Chiesura, A. (2004) The role of urban parks for the sustainable city *Landscape and Urban Planning* **68**: 129-138.

Choe, E., Lee, N., Lee, B., Pratt, W. and Kientz, J. (2014) Understanding quantified-selfers' practices in collecting and exploring personal data, In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 1143-1152. ACM.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J., Mellouli, S., Nahon, K. and Scholl, H. (2012) Understanding smart cities: An integrative framework, In: *System Science (HICSS), 2012 45th Hawaii International Conference on*, 2289-2297. IEEE.

Chung, S. and Liu, S. (2011) *Predicting stock market fluctuations from twitter*. Berkeley, California.

Clements, M., Serdyukov, P., De Vries, A. and Reinders, M. (2010) Using Flickr geotags to predict user travel behaviour, In: *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, 851-852. ACM.

Cohen, D., Ashwood, J., Scott, M., Overton, A., Evenson, K., Staten, L., Porter, D. McKenzie, T. and Catellier, D. (2006) Public parks and physical activity among adolescent girls *Pediatrics* **118 (5)**: e1381-e1389.

Cohen, D., Marsh, T., Williamson, S., Golinelli, D. and McKenzie, T. (2010) Impact and cost-effectiveness of family fitness zones: a natural experiment in urban public parks *Health & place* **18 (1)**: 39-45.

Cohen, D., McKenzie, T., Sehgal, A., Williamson, S., Golinelli, D. and Lurie, N. (2007) Contribution of public parks to physical activity *American Journal of Public Health* **97 (3)**: 509-514.

Cohen, D., Sehgal, A., Williamson, S., Marsh, T., Golinelli, D. and McKenzie, T. (2009a) New recreational facilities for the young and old: Policy and programming implications *Journal of Public Health Policy* **30 (1)**: S248-S263.

Coleman, D., Georgiadou, Y. and Labonte, J. (2009) Volunteered geographic information: The nature and motivation of producers *International Journal of Spatial Data Infrastructures Research* **4 (1)**: 332-358.

Collins, J., Kinzig, A., Grimm, N., Fagan, W., Hope, D., Wu, J. and Borer, E. (2000) A new urban ecology *American Scientist* **88**: 416-425.

Cooper, A., Page, A., Wheeler, B., Hillsdon, M., Griew, P. and Jago, R. (2010) Patterns of GPS measured time outdoors after school and objective physical activity in English children: the PEACH project *International Journal of Behavioral Nutrition and Physical Activity* **7 (1)**: 1.

Correa, T., Hinsley, A. and De Zuniga, H. (2010) Who interacts on the web?: the intersection of users' personality and social media use *Computers in Human Behaviour* **26 (2)**: 247-253.

Cosgrove, D. (1984) *Social Formation and Symbolic Landscape*. University of Wisconsin Press, United States.

Costanza, R., d'Arge, R., Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O' Neill, R., Paruelo, J., Raskin, R., Sutton, P. and Belt, M. (1997) The value of the world's ecosystem services and natural capital *Nature* **387**: 253–260.

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M. and Zegras, P. (2013) Future mobility survey: Experience in developing a smartphone-based travel survey in singapore *Transportation Research Record: Journal of the Transportation Research Board* **2354**: 59-67.

Coutts, A., Beringer, J. and Tapper, N. (2006) Impact of increasing urban density on local climate: spatial and temporal variations in the surface energy balance in Melbourne, Australia *Applied Meteorology and Climatolog*y **46**: 477-493.

Cresswell T (2004) *Place: A short introduction*. Wiley-Blackwell, New Jersey.

Daily, G. (1997) *Nature's services: societal dependence on natural ecosystems.* Island Press, Washington DC.

Dallimer, M., Tang, Z., Bibby, P., Brindley, P., Gaston, K. and Davies, Z. (2011) Temporal changes in greenspace in a highly urbanized region *Biology Letters*: rsbl20110025.

Daniel, T., Muhar, A., Arnberger, A., Aznar, O., Boyd, J., Chan, K., Costanza, R., Elmqvist, T., Flint, C., Goster, P., Gret-Regamey, A., Lave, R., Muhar, S., Pekner, M., Ribe, R., Schauppenlehner, T., Sikor, Soloviy, I., Spierenburg, M., Taczanowska, K., Tam, J. and von der Dunk, A. (2012) Contributions of cultural services to the ecosystem services agenda *Proceedings of the National Academy of Sciences* **109** (23): 8812-8819.

Daniel, T., Muhar, A., Arnberger, A., Aznar, O., Boyd, J., Chan, K., Costanza, R., Elmqvist, T., Flint, C., Gobster, P., Gret-Regamey, A., Lave, R., Muhar, S., Penker, M., Ribe, R., Schauppenlehner, T., Sikor, T., Soloviy, I., Spierenburg, M., Taczanowska, K., Tam, J. and von der Dunk, A. (2012) Contributions of cultural services to the services agenda *PNAS* **109 (23)**: 8812-8819.

Davidson J, Bondi L and Smith M (2005) *Emotional Geographies*. Ashgate Publishing, Aldershot.

Davidson, J. and Milligan, C. (2004) Embodying Emotion Sensing Space: Introducing emotional geographies *Social & Cultural Geographies* **5 (4)**: 523-532.

Davidson, R. and Sutton, S. (1995) Affective neuroscience: the emergence of a discipline *Current Opinion in Neurobiology* **5**: 217-224.

Davison, K. and Lawson, C. (2006) Do attributes in the physical environment influence children's physical activity? A review of the literature International *Journal of Behavioral Nutrition and Physical Activity* **3 (1)**: 1.

De Groot, R., Alkemade, R., Braat, L., Hein, L. and Willemen, L. (2010) Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making *Ecological complexity* **7 (3)**: 260-272.

De Groot, R., Ramakrishnan, P., Berg, A., Kulenthran, T., Muller, S., Pitt, D. and Wascher, D. (2005) *Ecosystem and human wellbeing: current state and trends, Volume 1, Findings of the Condition and Trends working Group of the Millennium Ecosystem Assessment Series*. Island Press, Washington DC.

De Valck, J., Broekx, S., Liekens, I., De Nocker, L., Van Orshoven, J. and Vranken, L. (2016) Contrasting collective preferences for outdoor recreation and substitutability of nature areas using hot spot mapping *Landscape and Urban Planning* **151**: 64-78.

Del Saz Salazar, S. and Garcia Menendez, L. (2007) Estimating the non-market benefits of an urban park: does proximity matter? *Land Use Policy* **24**: 296-305.

Deloitte (2015) *Mobile Consumer 2015: The UK Cut*. Deloitte, UK.

DeLyser, D. and Sui, D. (2012) Crossing the qualitative-quantitative divide II: Inventive approaches to big data, mobile methods, and rhythmanalysis *Progress in Human Geography* **37 (2)**: 293-305.

Demirbas, M., Bayir, M., Akcora, C., Yilmaz, Y. and Ferhatosmanoglu, H. (2010) Crowd-sourced sensing and collaboration using twitter, In: *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a*. 1-9. IEEE.

Department for Environment, Food and Rural Affairs (DEFRA) (2017) *Evidence statement on the links between natural environments and human health*. Department for Environment, Food and Rural Affairs and University of Exeter Medical School, United Kingdom.

Díaz, S. and Cabido, M. (2001) Vive la difference: plant functional diversity matters to ecosystem processes *Trends in ecology & evolution* **16** (11): 646-655.

Diener, E., Suh, E., Lucas, R. and Smith, H. (1999) Subjective well-being: Three decades of progress *Psychological Bulletin* **125 (2)**: 276-302.

Dijst, M. (2009) Time geographical analysis, In: Kitchin, R. and Thirft, N. [Eds] *International encyclopaedia of human geography*. 266-278. Oxford, Elsevier.

Dijst, M. (2013) Space–time integration in a dynamic urbanizing world: Current status and future prospects in geography and GIScience: Space–time integration in geography and GIScience *Annals of the Association of American Geographers* **103 (5)**: 1058-1061.

Dodds, P., Harris, K., Kloumann, I., Bliss, C. and Danforth, C. (2011) Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter *PLoS one* **6 (12)**: e26752.

Dyck, I. (1990) Space, time, and renegotiating motherhood: An exploration of the domestic workplace *Environment and Planning D* **8**: 459-83.

Edwards, N., Myer, G., Kalkwarf, H., Woo, J., Khoury, P., Hewett, T. and Daniels, S. (2015) Outdoor temperature, precipitation and wind speed affect physical activity levels in children: a longitudinal cohort study *Journal of Physical Activity & Health* **12 (8)**: 1074-1081.

Edwards, P. and Tsouros, A. (2006) *Promoting physical activity and active living in urban environments: the role of local government.* WHO Regional Office, Europe.

Egoh, B., Rouget, M., Reyers, B., Knight, A., Cowling, R., van Jaarsveld, A. and Welz, A. (2007) Integrating ecosystem services into conservation assessments: a review *Ecological Economics*, **63:** 714-721.

Ehrlich, P. and Ehrlich, A. (1981) *Extinction: the causes and consequences of the disappearance of species*. Random House, New York.

Ekman, P. (1992) An argument for basic emotions *Cognition & emotion* **6 (3-4)**: 169-200.

Ekman, P. (1999) Basic Emotions, In: Dalgleish, T. and Power, M. [eds] *Handbook of Cognition and Emotion*, 45-60. Wiley and Sons Ltd, New Jersey.

Ekman, P. and Friesen, W. (1971) Constants across cultures in the face and emotion *Journal of Personality and Social Psychology* **17 (2)**: 124.

Ellard, C. (2015) *Places of the heart The psychogeography of everyday life*. Bellevue Literary Press, New York.

Ellison, N. (2007) Social network sites: Definition, history, and scholarship *Journal of computer-mediated Communication* **13 (1):** 210-230.

Elmqvist, T. (2011) Introduction. Section 4 Ecosystem services, and social systems in Urban Landscapes, In: Niemela et al. [Eds] *Urban Ecology: Patterns, Processes and Applications*. 189-192. Oxford University Press, United Kingdom.

Elwood, S. (2008) Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS *GeoJournal* **72 (3-4)**: 173-183.

England, K. (1993) Suburban pink collar ghettos: The spatial entrapment of women? *Annals of the Association of American Geographers* **83**: 225-42.

Ergler, C., Kearns, R. and Witten, K. (2016) Exploring children's seasonal play to promote active lifestyles in Auckland, New Zealand *Heath and Place* **41** 67-77.

Ernstson, H., Sörlin, S. and Elmqvist, T. (2008) Social movements and ecosystem services—the role of social network structure in protecting and managing urban green areas in Stockholm *Ecology and Society*, **13 (2)**: 29.

Estellés-Arolas, E. and González-Ladrón-de-Guevara, F. (2012) Towards an integrated crowdsourcing definition *Journal of Information science* **38** (2): 189-200.

Evenson, K., Wen, F., Golinelli, D., Rodrigues, D. and Cohen, D. (2013) Measurement properties of a park use questionnaire *Environment and Behaviour* **45 (4)**: 526-547.

Evenson, K., Wen, F., Hillier, A. and Cohen, D. (2014) Assessing the contribution of parks to physical activity using GPS and accelerometry *Medicine and Science in Sports & Exercise* **45 (10)**: 1981-1987.

Ewert, A. and Heywood, J. (1991) Ground development in the natural environment. Expectations, outcomes and techniques *Environment and Behaviour* **23 (5)**: 592-615.

Farkas, J. (2014) *Shifting Gears. Part One: Smarter Cycling in Calgary*. Manning Foundation for Democratic Education, Calgary.

Feral, C. (1998) The connectedness model and optimal development: is ecopsychology the answer to emotional well-being? *The Humanistic Psychologist* **26**: 243-274.

Fisher, B., Turner, R. and Morling, P. (2009) Defining and classifying ecosystem services for decision making *Ecological Economics* **68**: 643–653.

Flanagin, A. and Metzger, M. (2008) The credibility of volunteered geographic information *GeoJournal* **72 (3-4)**: 137-148.

Floyd, M., Bocarro, J., Smith, W., Baran, P., Moore, R., Cosco, N., Edwards, M., Suau, L. and Fang, K. (2011) Park-based physical activity among children and adolescents *American Journal of Preventive Medicine* **41 (3)**: 258-265.

Forget, G. and Lebel, J. (2001) An ecosystem approach to human health *International Journal of Occupational and Environmental Health* **7 (2)**: 3-38.

Forman, R. and Godron, M. (1981) Patches and structural components for a landscape ecology *BioScience*, **31** (10): 733-740.

Frank, L. and Kavage, S. (2009) A national plan for physical activity: the enabling role of the built environment *Journal of Physical Activity and Health* **6 (2)**: s185-s196.

Frank, L., Schmid, T., Sallis, J., Chapman, J. and Saelens, B. (2005) Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ *American journal of preventive medicine* **28 (2)**: 117-125.

Frias-Martinez, V. and Frias-Martinez, E. (2014) Specteral clustering for sensing urban land use using Twitter activity *Engineering Applications of Artificial Intelligence* **35**: 237-245.

Friberg, T. (1993) *Everyday Life: Women's Adaptive Strategies in Time And Space. Translated by Madi Gray*. Lund University Press, Lund, Sweden.

Fuller, R. and Gaston, K. (2009) The scaling of green space coverage in European cities *Biology Letters* **5 (3)**: 352-355.

Gaiani, M., Martini, B. and Apollonio, F. (2017) A Framework for a Smart Cultural City, In: *New Activities For Cultural Heritage*, 217-227. Springer, Cham.

Galea, S. and Vlahov, D. (2005) Urban health: evidence, challenges, and directions *Annual Review of Public Health* **26:** 341-365.

García-Palomares, J., Gutiérrez, J. and Mínguez, C. (2015) Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS *Applied Geography* **63:** 408-417.

Gedron, M. (2010) Defining Emotion: A brief history *Emotion Review* **2 (4)**: 371-372.

Geurs, K. and Van Wee, B. (2004) Accessibility evaluation of land-use and transport strategies: review and research directions *Journal of Transport geography* **12 (2)**: 127-140.

Girardin, F., Vaccari, A., Gerber, A. Biderman, A. and Ratti, C. (2009) Quantifying urban attractiveness from the distribution and density of digital footprints *International Journal of Spatial Data Infrastructures Research* **4**: 175-200.

Go, A., Bhayani, R. and Huang, L. (2009) Twitter sentiment classification using distant supervision *CS224N Project Report, Stanford* **1 (12)**.

Gobster, P., Nassauer, J., Daniel, T. and Fry, G. (2007) The shared landscape: what does aesthetics have to do with ecology? *Landscape Ecology* **22:** 959-972.

Gómez-Baggethun, E. and Barton, D. (2013) Classifying and valuing ecosystem services for urban planning *Ecological Economics* **86**: 235-245.

Gordon-Larsen, P., McMurray, R. and Popkin, B. (2000) Determinants of adolescent physical activity and inactivity patterns *Pediatrics* **105 (6)**: e83-e83.

Government Office for Science (GOS) (2007) *Tackling obesity: future choices – obesity systems atlas Government Office for Science.* Government Office for Science, United Kingdom.

Greenspace (2007) *The Park Life Report: The First Ever Public Satisfaction Survey of Britain's Parks and Green Spaces*. Greenspace, Berkshire.

Greider, T. and Garkovich, L. (1994) Landscapes: The social construction of nature and the environment *Rural sociology* **59 (1)**: 1-24.

Griffin, G. and Jiao, J. (2015) Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus *Journal of Transport & Health* **2 (2)**: 238-247.

Grimm N., Grove J., Pickett S. and Redman C. (2000) Integrated approaches to long term studies of urban ecological systems *BioScience* **50**: 571–584.

Groenewegen, P., Berg, A., Vries, S. and Verheij, R. (2006) Vitamin G: effects of green space on health, well-being, and social safety *BMC public health* **6 (1)**: 1.

Grossman, L. (1977) Man-environment relationships in anthropology and geography *Annals of the Association of American Geographers* **67** (1): 126-144.

Grove, J. and Burch, W. (1997) A social ecology approach and applications of urban ecosystem and landscape analyses: a case study of Baltimore, Maryland *Urban Ecosystems* **1** (4): 259-275.

Grzywacz, J.and Fuqua, J. (2000) The social ecology of health: leverage points and linkages *Behavioral Medicine* **26** (3): 101-115.

Guardian (2017) *How green is your city? UK's top 10 mapped and ranked* https://www.theguardian.com/cities/gallery/2017/jan/05/green-space-uk-largest-cities-mapped#img-6 Accessed 11/10/2017.

Guba, E. and Lincoln, Y. (1998) Competing paradigms in qualitative research, In: Denzin, N. and Lincoln, Y. [Eds] *The landscape of qualitative research: Theories and issues*, 195-220. Thousand Oaks, Sage, California.

Guerrero , P., Moller, M., Olafsson, A. and Snizek, B. (2016) Revealing cultural ecosystem services through Instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance *Urban Planning* **1 (2)**: 1-17.

Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., Gomez-Baggethun, E., Gren, A., Hamstead, Z., Hansen, R., Kabisch, N., Kremer, P., Langemeyer, J., Rall, E. and McPhearson, T. (2014) A quantitative review of urban ecosystem service assessments: concepts, models, and implementation *Ambio* **43 (4)**: 413-433.

Haines-Young, R. and Potschin, M. (2010) The links between biodiversity, ecosystem services and human well-being, In: Raffaelli, D. and Frid, C. [eds] *Ecosystem Ecology: a new synthesis BES Ecological Reviews Series*, 110-139. Cambridge University Press, Cambridge.

Hale, J. (2015) *Elucidating the drivers, contextual sensitivity and resilience of urban ecological systems* (PhD Thesis). University of Birmingham, United Kingdom.

Han, B., Cohen, D. and McKenzie, T. (2013) Quantifying the contribution of neighborhood parks to physical activity *Preventive medicine* **57 (5)**: 483-487.

Hancke, G., de Carvalho e Silva, B. and Hancke Jnr, G. (2013) The role of advanced sensing in smart cities *Sensors* **13 (1)**: 393-425.

Handley, J., Pauleit, S., Slinn, P., Lindley, S., Baker, M., Barber, A. and Jones, C. (2003) *Providing accessible natural green space in towns and cities: a practical guide to assessing the resource and implementing local standards for provision*. Countryside Council for Wales Nature England, United Kingdom.

Hannay, P. and Baatard, G. (2011) GeoIntelligence: Data mining locational social media content for profiling and information gathering, In: *Proceedings of the 2nd International Cyber Resilience Conference.* Australia.

Hanson, S. (2005) Perspectives on the geographic stability and mobility of people in cities *Proceedings of the National Academy of Sciences of the United States of America* **102 (43)**: 15301-15306.

Hartig, T. and Staats, H. (2006) The need for psychological restoration as a determinant of environmental preferences *Journal of Environmental Psychology* **26 (3)**: 215-226.

Hasan, S., Schneider, C., Ukkusuri, S. and González, M. (2013). Spatiotemporal patterns of urban human mobility *Journal of Statistical Physics* **151 (1-2)**: 304-318.

Hauthal, E. and Burghardt, D. (2013) Extraction of location-based emotions from photo platforms, In: Krisp, J. [Ed] *Progress in location-based services.* Springer, United States.

Hawker, M. (2010) *Developer's Guide to Social Programming: Building Social Context Using Facebook, Google Friend Connect, and the Twitter API*. Pearson Education, United Kingdom.

Health and Social Care Information Centre (HSCIC) 2015 *Statistics on Obesity, Physical Activity and Diet: England 2015.* Health and Social Care Information Centre, United Kingdom.

Herzog, T. and Chernick, K. (2000) Tranquility and danger in urban and natural settings *Journal of Environmental Psychology* **20 (1)**: 29-39.

Highfield, T. and Leaver, T. (2014) A methodology for mapping Instagram hashtags *First Monday* **20**: 1-5.

Hitchings, R. (2010) Seasonal climate change and the indoor city worker *Transactions of the Institute of British Geographers* **35 (2)**: 282-298.

Hoebeke, R. (2008) Low-income women's perceived barriers to physical activity: focus group results *Applied Nursing Research* **21 (2)**: 60-65.

Hoehner, C., Ramirez, L., Elliott, M., Handy, S. and Brownson, R. (2005) Perceived and objective environmental measures and physical activity among urban adults *American Journal of Preventive Medicine* **28 (2)**: 105-116.

Hollander, J., Graves, E., Renski, H., Foster-Karim, C., Wiley, A. and Das, D. (2016) *Urban social listening: Potential and pitfalls for using microblogging data in studying cities*. Palgrave Macmillan, United Kingdom.

Honold, J., Lakes, T., Beyer, R. and van der Meer, E. (2016) Restoration in Urban Spaces Nature Views From Home, Greenways, and Public Parks *Environment and Behavior* **48 (6)**: 796-825.

Howe, P. and Morris, C. (2009) An exploration of the co-production of performance running bodies and natures within “running taskscapes” *Journal of sport and social issues* **33 (3)**: 308-330.

HSCIC (2016) *Statistics on obesity, physical activity and diet, England 2016.* HSCIC, United Kingdom.

Huberman, B., Romero, D. and Wu, F. (2008) Social networks that matter: Twitter under the microscope *First Monday* **14**: 1-9.

Hug, S., Hartig, T., Hansmann, R., Seeland, K. and Hornung, R. (2009) Restorative qualities of indoor and outdoor exercise settings as predictors of exercise frequency *Health and Place* **15 (4)**: 971-980.

Humpel, N., Owen, N. and Leslie, E. (2002) Environmental factors associated with adults’ participation in physical activity: a review *American Journal of Preventative Medicine* **22 (3)**: 188-199.

Illingworth, S., Muller, C., Graves, R. and Chapman, L. (2014) UK Citizen Rainfall Network: a pilot study *Weather* **69 (8)**: 203-207.

Ingold, T. (2000) *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*. London, Routledge.

Internet Live Stats (2017) *Twitter Usage Statistics*, http://www.internetlivestats.com/twitter-statistics/#ref-3 Accessed 28/04/2017.

Irwin, E. and Bockstael, N. (2007) The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation *Proceedings of the National Academy of Sciences* **104 (52)**: 20672-20677.

Jansen, B., Zhang, M., Sobel, K. and Chowdury, A. (2009) Twitter power: Tweets as electronic word of mouth *Journal of the American society for information science and technology* **60 (11)**: 2169-2188.

Java, A., Song, X., Finin, T. and Tseng, B. (2007) Why we twitter: understanding microblogging usage and communities *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 56-65. ACM.

Johnson, G., Bagstad, K., Snapp, R. and Villa, F. (2012) Service Path Attribution Networks (SPANs): A Network Flow Approach to Ecosystem Service Assessment *International Journal of Agricultural and Environmental Information Systems* **3 (2)**: 54-71.

Jones, A., Hillsdon, M. and Coombes, E. (2009) Greenspace access, use, and physical activity: understanding the effects of area deprivation *Preventative Medicine* **49 (6)**: 500- 505.

Jones, P. and Evans, J. (2012) Rescue Geography: Place Making, Affect and Regeneration *Urban Studies* **49 (11)**: 2315-2330.

Joseph, R. and Maddock, J. (2016) Observational Park-based physical activity studies: A systematic review of the literature *Preventive Medicine* **89**: 257-277.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M. and Newth, D. (2015) Understanding human mobility from Twitter *PloS one* **10 (7)**: e0131469.

Kabisch, N., Qureshi, S. and Haase, D. (2015) Human-environment interactions in urban green space – A systematic review of contemporary issues and prospects for future research *Environmental Impact Assessment Review* **50**: 25-34.

Kaczynski, A., Potwarka, L. and Saelens, B. (2008) Association of park size, distance, and features with physical activity in neighborhood parks *American Journal of Public Health* **98 (8)**: 1451-1456.

Kaczynski, A., Potwarka, L., Smale, B. and Havitz, M. (2009) Association of parkland proximity with neighborhood and park-based physical activity: variations by gender and age *Leisure Sciences* **31 (2)**: 174-191.

Kahn, P. (1999) *The human relationship with nature: Development and culture*. MIT Press, Massachusetts.

Kalin, J. and Frith J. (2016) Wearing the city: memory p(a)laces, smartphones, and the rhetorical invention of embodied space *Rhetoric Society Quarterly* **46 (3)**: 1-14.

Kallus, R. (2010) From abstract to concrete: subjective reading of urban space *Urban Design* **6 (2)**: 129-150.

Kanhere, S. (2011) Participatory sensing: crowdsourcing data from mobile smartphones in urban spaces, In: *Mobile Data Management (MDM), 2011 12th IEEE International Conference.* IEEE.

Kaplan, R. (2001) The nature of the view from home: psychological benefits *Environment & Behavior* **33**: 507-542.

Kaplan, A. and Haenlein, M. (2010) Users of the world, unite! The challenges and opportunities of Social Media *Business horizons* **53 (1):** 59-68.

Kellert, S. and Wilson, E. (1995) *The biophilia hypothesis*. Island Press, Washington DC.

Keniger, L., Gaston, K., Irvine, K. and Fuller, R. (2013) What are the benefits of interacting with nature? *International journal of environmental research and public health* **10 (3)**: 913-935.

Kim, J. and Kaplan, R. (2004) Physical and psychological factors in sense of community new urbanist Kentlands and nearby Orchard Village *Environment and behavior* **36 (3)**: 313-340.

Kitchin, R (2014) The real-time city? Big data and smart urbanism *Geojournal* **79 (1)**: 1-14.

Kittinger, J., Coontz, K., Yuan, Z., Han, D., Zhao, X. and Wilcox, B. (2009) Toward holistic evaluation and assessment: linking ecosystems and human well-being for the Three Gorges Dam *EcoHealth* **6 (4)**: 601-613.

Kleeman, F., Voss, G. and Rieder, K. (2008) Un(der)paid innovators: the commercial utilization of consumer work through crowdsourcing *Science, Technology and Innovation Studies* **4 (1)**: 5-26.

Klettner, S., Huang, H., Schmidt, M. and Gartner, G. (2013) Crowdsourcing affective responses to space *Kartographische Nachrichten* **2 (3)**: 66-72.

Korpela, K. (2003) Negative mood and adult place preference *Environment and Behavior* **35 (3)**: 331-346.

Kövecses, Z. (2003) *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge, Cambridge University Press.

Kremen, C. (2005) Managing ecosystem services: what do we need to know about their ecology? *Ecology letters* **8** (5): 468-479.

Kremer, P., Hamstead, Z., Haase, D., McPhearson, T., Frantzeskaki, N., Andersson, E., Kabisch, N., Larondelle, N., Rall, E., Voigt, A., Baro, F., Bertram, C., Gomez-Baggethan, E., Hansen, R., Kaczorowska, A., Kain, J., Kronenberg, J., Mangemeyer, J., Pauleit, S., Rehdanz, K., Schewenius, M., van Ham, C., Wurster, D. and Elmqvist, T. (2016) Key insights for the future of urban ecosystem services research *Ecology and Society* **21 (2)**: 29.

Krenichyn, K. (2006) 'The only place to go and be in the city': women talk about exercise, being outdoors, and the meanings of a large urban park *Health & Place* **12 (4)**: 631-643.

Krenn, P., Oja, P. and Titze, S. (2014) Route choices of transport bicyclists: a comparison of actually used and shortest routes *International journal of behavioral nutrition and physical activity* **11 (1)**: 31.

Krenn, P., Titze, S., Oja, P., Jones, A. and Ogilvie, D. (2011) Use of global positioning systems to study physical activity and the environment: a systematic review *American journal of preventive medicine* **41 (5)**: 508-515.

Kwak, H., Lee, C., Park, H. and Moon, S. (2010) What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web,* 591-600. ACM.

Kwan, M. (2004) GIS methods in time-geographic research: geocomputation and geovisualisation of human activity patterns *Geografiska Annaler: Series B, Human Geography* **86B (4)**: 267-280.

Kwan, M. and Lee, J. (2003) Geovisualisation of Human Activity Patterns Using 3D GIS: A time-geographic approach, In: Goodchild, M. and Janelle, D. [Eds] *Spatially integrated social science: Examples and Best Practice,* 1-23. Oxford University Press, Oxford.

Kweon, B., Sullivan, W. and Wiley, A. (1998) Green common spaces and the social integration of inner-city older adults *Environment and behavior* **30 (6)**: 832-858.

Lafortezza, R., Carrus, G., Sanesi, G., and Davies, C. (2009) Benefits and well-being perceived by people visiting green spaces in periods of heat stress *Urban Forestry & Urban Greening* **8 (2)**: 97-108.

Landis, J. and Koch, G. (1977) The measurement of observer agreement for categorical data *Biometrics* **33 (1)**: 159–174.

Lane, R. and Nadel, L. (2002) *Cognitive Neuroscience of emotion*. Oxford University Press, Oxford.

Lange, P. (2007) Publicly private and privately public: Social networking on YouTube *Journal of computer‐Mediated communication* **13 (1):** 361-380.

Larondelle, N., Haase, D. and Kabisch, N. (2014) Mapping the diversity of regulating ecosystem services in European cities *Global Environmental Change* **26**: 119-129.

LeDantec, C., Asad, M., Misra, A. and Watkins, K. (2015) Planning with crowdsourced data: Rhetoric and representation in transportation planning, In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1717-1727. ACM.

LeDoux, J. (1996) *The emotional brain*. Simon & Schuster, New York.

Lee, A. and Maheswaran, R. (2010) The health benefits of urban green spaces: a review of the evidence *Journal of Public Health* fdq068.

Lee, Y. (2005) Gender differences in physical activity and walking among older adults Journal of *Women & Aging* **17 (1-2)**: 55-70.

Lenhart, A., Purcell, K., Smith, A. and Zickuhr, K. (2010) *Social media & mobile internet use among teens and young adults Millennials*. Pew Internet & American Life Project, United States.

Lenntrop, B. (1999) Time geography at the end of its beginning *Geojournal* **48**: 155-158.

Liu, B. (2012) Sentiment analysis and opinion mining *Synthesis lectures on human language technologies* **5 (1)**: 1-167.

Liu, J., Dietz, T., Carpenter, S., Alberti, M., Folke, C., Moran, E., Pell, A., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C., Schneider, S. and Taylor, W. (2007) Complexity of coupled human and natural systems *Science* **317**: 1513-1516.

Lockwood, M. and Tracey, K. (1995) Nonmarket economic valuation of an urban green park *Journal of Leisure Research* **27 (2)**: 155-167.

Low, S., Taplin, D., and Scheld, S. (2009) *Rethinking urban parks: Public space and cultural diversity*. University of Texas Press, Texas.

Lutz, C. and White, G. (1986) The anthropology of emotions *Annual Review of Anthropology* **15**: 405-436.

Ma, X., Wei, Z., Chai, Y. and Xie, K. (2008) Intergrating map services and location-based services for georeferenced individual data collection, In: *Geoscience and Remote Sensing Symposium, 2008 IGARSS 2008 IEEE International* 5. IEEE, V-81.

Ma, Y., Olendzki, B., Li, W., Hafner, A., Chiriboga, D., Hebert, J., Campbell, M., Sarnie, M. and Ockene, I. (2006) Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population *European Journal of Clinical Nutrition* **60 (4)**: 519-528.

Maas, J., van Dillen, S. and Verheij, R. (2009) Social contacts as a possible mechanism behind the relation between green space and health *Health Place* **15 (2)**: 586-595.

Maas, J., Verheij, R., Groenewegen, P., De Vries, S. and Spreeuwenberg, P. (2006) Green space, urbanity, and health: how strong is the relation? *Journal of epidemiology and community health* **60** (7): 587-592.

Maas, J., Verheij, R., Spreeuwenberg, P. and Groenewegen, P. (2008) Physical activity as a possible mechanism behind the relationship between green space and health: a multilevel analysis *BMC Public Health* **8 (1)**: 1.

MacKerron, G. and Mourato, S. (2013) Happiness is greater in natural environments *Global Environmental Change* **23 (5)**: 992-1000.

Madden, M., Lenhart, A., Duggan, M., Cortesi, S. and Gasser, U. (2013) *Teens and technology 2013*. Pew Internet & American Life Project, Washington, DC.

Maller, C., Townsend, M., Pryor, A., Brown, P. and St Leger, L. (2006) Healthy nature healthy people: ‘contact with nature’ as an upstream health promotion intervention for populations *Health promotion international* **21 (1):** 45-54.

Malleson, N. and Andresen, M. (2015) The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns *Cartography and Geographic Information Science* **42 (2)**: 112-121.

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B. and Rodrigue, J. (2012) A demographic analysis of online sentiment during hurricane Irene, In: *Proceedings of the Second Workshop on Language in Social Media,* 27-36. Association for Computational Linguistics.

Martınez, V. and González, V. (2013) Sentiment characterization of an urban environment via Twitter, In: *Ubiquitous computing and ambient intelligence. Context-awareness and context-driven interaction,* 394-397. Springer Verlag, Berlin.

Marwick, A. and Boyd, D. (2011) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* **13 (1):** 114-133.

Mathioudakis, M. and Koudas, N. (2010) Twittermonitor: trend detection over the twitter stream *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, 1155-1158. ACM, Indianapolis.

Matthews, C., Freedson, P., Herbert, J., Stanek, E., Merriam, P., Rosal, M., Ebbeling, C. and Ockene, I. (2001) Seasonal variation in household, occupational and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study *American Journal of Epidemiology* **153**: 172-183.

Mayer, F., Frantz, C., Bruehlman-Senecal, E. and Dolliver, K. (2008) Why is nature beneficial? The role of connectedness to nature *Environment and Behavior* **41 (5)**: 607-643.

McCormack, G., Rock, M., Toohey, A. and Hignell, D. (2010) Characteristics of urban parks associated with park use and physical activity: A review of qualitative research *Health & Place* **16 (4)**: 712-726.

McHugh, M. (2012) Interrater reliability: the kappa statistic, *Biochem Med* **22 (3)**: 276-282.

McIntyre, N., Knowles-Yanez, K. and Hope, D. (2000) Urban ecology as an interdisciplinary field: differences in the use of ‘urban’ between the social and natural sciences *Urban Ecosystems* **4**: 5-24.

McKenzie, T., Cohen, D., Sehgal, A., Williamson, S. and Golinelli, D. (2006) System for observing play and recreation in communities (SOPARC): reliability and feasibility measures *Journal of Physical Activity and Health* **6 (1)**: S208-S222.

McLaren, L. and Hawe, P. (2005) Ecological perspectives in health research *Journal of Epidemiology and Community Health* **59**: 6-14

McLeroy, K., Bibeau, D., Steckler, A. and Glanz, K. (1988) An ecological perspective on health promotion programs *Health Education & Behavior* **15** (4): 351-377.

McMahan, E. and Estes, D. (2015) The effect of contact with natural environments on positive and negative affect: A meta-analysis *The Journal of Positive Psychology* **10 (6)**: 507-519.

Mehta, M. and Ouellet, E. (1995) *Environmental Sociology: Theory and Practice*. Captus Press, Ontario.

Meijer, A. and Bolívar, M. (2016) Governing the smart city: a review of the literature on smart urban governance *International Review of Administrative Sciences* **82 (2)**: 392-408.

Merchant, A., Dehghan, M. and Akhtar-Danesh, N. (2007) Seasonal variation in leisuretime physical activity among Canadians *Canadian Journal of Public Health* **98 (3)**: 203-208.

Midgley, P. (2011) Bicycle-sharing schemes: enhancing sustainable mobility in urban areas, United Nations *Department of Economic and Social Affairs*: 1-12.

Milcu, A., Hanspach, J., Abson, D. and Fischer, J. (2013) Cultural ecosystem services: A literature review and prospects for future research *Ecology and Society*, **18** (3): 44.

Millennium Ecosystem Assessment (MEA) (2005) *Ecosystems and human well-being.* **5** Island Press, Washington DC.

Miller, J. (2005) Biodiversity conservation and the extinction of experience *Trends in Ecology and Evolution* **20:** 430-434.

Millward, A. and Sabir, S. (2011) Benefits of a forested urban park: What is the value of Allan Gardens to the city of Toronto, Canada? *Landscape and urban planning* **100 (3)**: 177-188.

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J. and Rosenquist, J. (2011) Understanding the demographics of twitter users. *ICWSM* **11**: 5.

Mitchell, D. (1995) The end of public space? People's Park, definitions of the public, and democracy *Annals of the Association of American Geographers* **85 (1)**: 108-133.

Mitchell, L., Frank, M., Harris, K., Dodds, P. and Danforth, C. (2013) The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place *PLoS one* **8 (5)**: e64417.

Mitchell, R., and Popham, F. (2007) Greenspace, urbanity and health: relationships in England *Journal of Epidemiology and community health* **61** (8): 681-683.

Moore, R. and Young, D. (1978) Childhood outdoors: Toward a social ecology of the landscape, In: *Children and the environment*, 83-130. Springer, United States.

Morstatter, F., Pfeffer, J., Liu, H. and Carley, K. (2013) Is the sample good enough? Comparing data from twitter's streaming API with twitter's firehose, In: *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* Massachusetts.

Muller, C., Chapman, L., Grimmond, C., Young, D. and Cai, X. (2013) Sensors and the city: a review of urban meteorological *networks International Journal of Climatology* **33 (7)**: 1585-1600.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential *International Journal of Climatology* **35**: 3185-3203.

Muñoz, S. (2009) *Children in the outdoors: a literature review.* Sustainable Development Research Centre, Edinburgh.

Murthy, D., Gross, A. and Pensavelle, A. (2015) Urban social media demographics: An exploration of Twitter use in major American cities *Journal of computer-mediated* communication **21** (1): 33-49.

Musakwa, W. and Selala, K. (2016) Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg, South Africa *Data in brief* **9**: 898-905.

Naaman, M., Boase, J. and Lai, C. (2010) Is it Really About Me? Message Content in Social Awareness Streams, In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work, 1*89-192. ACM.

National Health Service (NHS) (2016) *Inpatients formally detained in hospitals under the Mental Health Act 1983 and patients subject to Supervised Community Treatment: 2015/16, Annual Figures*. NHS Digital, United Kingdom.

Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G. and Scorrano, F. (2014) Current trends in Smart City initiatives: Some stylised facts *Cities* **38**: 25-36.

Nguyen, Q., Kath, S., Meng, H., Li, D., Smith, K., VanDerslice, J., Wen, M. and Li, F. (2016) Leveraging geotagged Twitter data to examine neighbourhood happiness, diet and physical activity *Applied Geography* **73**: 77-88.

Niemela, J. Saarela, S., Soderman, T., Kopperoinen, L., Yli-Pelkonen, V., Vare, S. and Kotze, J. (2010) Using the ecosystem service approach for better planning and conservation of urban green spaces: a Finland case study *Biodiversity and Conservation* **19 (11)**: 3225-3243.

Nikolopoulou, M., Baker, N. and Steemers, K. (2001) Thermal comfort in outdoor urban spaces: understanding the human parameter *Solar Energy* **70**: 227–235.

Nisbet, E. and Zelenski, J. (2011) Underestimating nearby nature affective forecasting errors obscure the happy path to sustainability *Psychological science* **22 (9)**: 1101-1106.

Noulas, A., Scellato, S., Mascolo, C. and Pontil, M. (2011) Exploiting semantic annotations for clustering geographic areas and users in location-based social networks, In: *Proceedings of the International Conference on Weblogs and Social Media. ACM.*

Ochsner, K. and Gross, J. (2007) The neural architecture of emotion regulation *Handbook of Emotion Regulation* **1 (1)**: 87-109.

Odland, J. (1998) Longitudinal analysis of migration and mobility spatial behavior in explicitly temporal contexts, In: Egenhofer, M. and Golledge, R. [Eds] *Spatial and Temporal Reasoning in Geographic Information Systems*, 238-260. Oxford University Press, Oxford.

Odum, E. (1971) *Fundamentals of Ecology*. W. B. Saunders Company, Philadelphia.

Ofcom (2015) *The Communications Market Report*. Ofcom, United Kingdom.

Office for National Statistics (ONS) (2016*) Statistics on Obesity, Physical Activity and Diet – England*. Office for National Statistics, United Kingdom.

Office for National Statistics (ONS) *2014 Mid-2014 Mid-year population estimates BDB2015/04 Birmingham Demographic Briefing*. Office for National Statistics, United Kingdom.

Office for National Statistics (2013) 2011 *Census Analysis – Comparing rural and urban areas of England and Wales*. Office for National Statistics, United Kingdom.

Orr, S., Paskins, J. and Chaytor, S. (2014*) Valuing Urban Green Space: Challenges and Opportunities UCL Policy Briefing.* UCL Public Policy, London.

Overeem, A., Robinson, J., Leijnse, H., Steeneveld, G., Horn, B. and Uijlenhoet, R. (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures *Geophysical Research Letters* **40 (15)**: 4081-4085.

Pak, A. and Paroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining *LREC*, **10**.

Patterson, M. and Williams, D. (2005) Maintaining research traditions on place: Diversity of thought and scientific progress *Journal of Environmental Psychology* **25**: 361-380.

Pavalanathan, U. and Eisenstein, J. (2015) Emoticons vs. emojis on Twitter: A causal inference approach *arXiv preprint arXiv:1510.08480.*

Penedo, F. and Dahn, J. (2005) Exercise and well-being; a review of mental and physical health benefits associated with physical activity *Current Opinion in Psychiatry* **18 (2)**: 189-193.

Peng, J., Chen, X., Liu, Y., Lu, H. and Hu, X. (2016) Spatial identification if multifunctional landscapes and associated influencing factors in the Beijing-Tianjin-Hebei region, China *Applied Geography* **74**: 170-181.

Pert, C. (1997) *Molecules of emotion: Why you feel the way you feel*. Simon and Schuster, New York.

Pettit, C., Widjaja, I., Russo, P., Sinnot, R., Stimson, R. and Tomko, M. (2012) Visualisation support for exploring urban space and place, In: *XXII ISPRS Congress, Technical Commission IV* (Vol. 25).

Public Health England (PHE) (2017) *Spatial planning for health: an evidence resource for planning and designing healthier places*. PHE, United Kingdom.

Phithakkitnukoon, S. and Oliver, P. (2011) Sensing urban social geography using online social networking data, In: *Proceedings of the International Conference on Weblogs and Social Media.*

Pickett, S., Burch, W., Dalton, S., Foresman, T., Grove, J. and Rowntree, R. (1997) A conceptual framework for the study of human ecosystems in urban areas *Urban ecosystems*, **1** (4): 185-199.

Pickett, S., Cadenasso, M. and Grove, J. (2004) Resilient cities: Meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms *Landscape and Urban Planning* **69**: 369–384.

Pivarnik, J., Reeves, M. and Rafferty, A. (2003) Seasonal variation in adult leisure-time physical activity *Medicine and Science in Sports and Exercise* **35 (6)**: 1004-1008.

Plieninger, T., Dijks, S., Oteros-Rozas, E. and Bieling, C. (2013) Assessing, mapping, and quantifying cultural ecosystem services at community level *Land use policy* **33**: 118-129.

Postolache, O., Dias Pereira, J. and Girao, P. (2009) Smart Sensor Network for Air Quality Monitoring Applications *IEEE Transactions on Instrumentation and Measurement* **58 (9)**: 3253-3262.

Quercia, D., Ellis, J., Capra, L. and Crowcroft, J. (2012) Tracking gross community happiness from tweets, In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, 965-968. ACM.

Raento, M., Oulasvirta, A. and Eagle, N. (2009) Smartphones: An emerging tool for social scientists *Sociological Methods & Research* **37** (3): 426-454.

Rapport, D., Howard, J., Lannigan, R. and McCauley, W. (2003) Linking health and ecology in the medical curriculum *Environment International* **29** (2-3): 353-358.

Ratti, C., Frenchman, D., Pulselli, R. M. and 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.

Redman, C., Grove, J. and Kuby, L. (2004) Integrating social science into the long-term ecological research (LTER) network: social dimensions of ecological change and ecological dimensions of social change *Ecosystems* **7 (2)**: 161-171.

Resch, B. (2013) People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements, In: Krisp, J. [Ed] *Progress in Location-Based Services*, 391-406. Springer, Berlin Heidelberg.

Resch, B., Summa, A., Sagl, G., Zeile, P. and Exner, J. (2015) Urban emotions—Geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data, In: Gartner, G. and Huang, H. [Eds] *Progress in Location-Based Services 2014*, 199-212. Springer International Publishing, Switzerland.

Resch, B., Summa, A., Zeile, P. and Strube, M. (2016) Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary space-time linguistic algorithm *Urban Planning* **1 (2)**: 114-127.

Rishbeth, C. and Finney, N. (2006) Novelty and nostalgia in urban greenspace: refugee perspectives *Tijdschrift voor economische en sociale geografie* **97 (3)**: 281-295.

Roberts, H. (2017) Using Twitter data in urban green space research: A case study and critical evaluation *Applied Geography* **81**: 13-20.

Roberts, H., Resch, B., Sadler, J., Chapman, L. and Petutschnig, A. (in press) Investigating the emotional responses of individuals to urban green space using Twitter data: a critical comparison of three different methods of sentiment analysis *Urban Planning*

Roberts, H., Sadler, J. and Chapman, L. (2017) Using Twitter to investigate seasonal variation in physical activity in urban green space *GEO: Geography and Environment* **4 (2)**: e00041

Roberts, H., Sadler, J. and Chapman, L. (in press) The value of Twitter data in determining the emotional responses of people to urban green spaces: a case study and critical evaluation *Urban Studies*

Roberts, K., Roach, M., Johnson, J., Guthrie, J. and Harabagiu, S. (2012) EmpaTweet: Annotating and Detecting Emotions on Twitter *LREC:* 3806-3813.

Rodaway, P. (1994) *Sensuous Geographies*: *Body, Sense and Place*. Routledge, London

Rosenthal, S., Ritter, A., Nakov, P. and Stoyanov, V. (2014) Semeval-2014 task 9: Sentiment analysis in twitter, In: *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014),* 73-80.

Russell, M. (2013) *Mining the Social Web: Data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media Inc, California.

Saegert, S. and Winkel, G. (1990) Environmental Psychology *Annual Review of Psychology* **41:** 441-744.

Sagl, G., Resch, B., Hawelka, ,B. and Beinat, E. (2012) From social sensor data to collective human behaviour patterns – analysing and visualising spatio-temporal dynamics in urban environments, In: Jekel, T., Car, A., Strobl, J. and Griesebner, G. [Eds] *Proceedings of the GI-Forum,* 54-63.

Saif, H., Fernez, M., He, Y. and Alani, H. (2013) Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold, In: *1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013).* Turin, Italy.

Sallis, J. and Saelens, B. (2000) Assessment of physical activity by self report: status, limitations, and future directions *Research Quarterly for Exercise and Sport* **71**: 409.

Sallis, J., Bull, F., Burdett, R., Frank, L., Griffiths, P., Giles-Corti, B. and Stevenson, M. (2016). Use of science to guide city planning policy and practice: how to achieve healthy and sustainable future cities *The lancet*, **388 (10062)**, 2936-2947.

Sallis, J., Bauman, A. and Pratt, M. (1998) Environmental and policy interventions to promote physical activity *American Journal of Preventive Medicine* **15 (4)**: 379-397.

Sallis, J., Hovell, M., Hofstetter, C., Elder, J., Hackley, M., Caspersen, C. and Powell, K. (1990) Distance between homes and exercise facilities related to frequency of exercise among San Diego residents *Public Health Reports* **105 (2)**: 179.

Salmon, J., Owen, N., Crawford, D., Bauman, A. and Sallis, J. (2003) Physical activity and sedentary behaviour: A population-based study of barriers, enjoyment and preference *Health Psychology* **22**: 178-188.

Sauer, C. (1925) The morphology of landscape *University of California Publications in Geography* **2 (2)**: 29-54.

Scheer, M. (2012) Are emotions a kind of practice (and is that what makes them have a history)? A Bourdieuian approach to understanding emotion *History and Theory* **51**: 193-220.

Schipperijn, J., Sttigsdotter, U., Randrup, T. and Troelsen, J. (2010) Influences on the use of urban green space – A case study in Odense, Denmark *Urban Forestry & Urban Greening* **9 (1)**: 25-32.

Scholes, R., Reyers, B., Biggs, R., Spierenburg, M. and Duriappah, A. (2013) Multi-scale and cross-scale assessments of social–ecological systems and their ecosystem services *Current Opinion in Environmental Sustainability* **5 (1)**: 16-25.

Schwartz, R. and Hochman, N. (2014) *The social media life of public spaces: reading places through the lens of geo-tagged data.* Locative Media Routledge, New York.

South Downs National Park Authority (SDNPA) (2013) *Partnership Management Plan,* SDNPA, England.

Seeland, K., Dübendorfer, S. and Hansmann, R. (2009) Making friends in Zurich's urban forests and parks: the role of public green space for social inclusion of youths from different cultures *Forest Policy and Economics* **11 (1)**: 10-17.

Seto, K., Fragkias, M., Guneralp, B. and Reilly, M. (2011) A meta-analysis of global urban land expansion *PLOS one* **6** (8): e23777.

Shalunts, G., Backfried, G. and Prinz, P. (2014) Sentiment analysis of German social media data for natural disasters, In: *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM).*

Shamai, S. and Ilatov, Z. (2005) Measuring sense of place: Methodological aspects *Tijdschrift voor economische en sociale geografie* **96 (5)**: 467-476.

Shanahan, D., Franco, L., Lin, B., Gaston, K.and Fuller, R. (2016) The Benefits of Natural Environments for Physical Activity *Sports medicine* **46 (7)**: 989–95.

Shanahan, D., Fuller, R., Bush, R., Lin, B. and Gaston, K. (2015) The health benefits of urban nature: how much do we need? *Bioscience:* biv032.

Shanahan, D., Lin, B., Gaston, K., Bush, R. and Fuller, R. (2014) What is the role of trees and remnant vegetation in attracting people to urban parks? *Landscape Ecology* **30 (1): 153-165.**

Shelton, T., Poorthius, A. and 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.

Sherrouse, B., Clement, J. and Semmens, D. (2011) A GIS application for assessing, mapping, and quantifying the social values of ecosystem services *Applied geography* **31 (2)**: 748-760.

Shinew, K., Glover, T. and Parry, D. (2004) Leisure spaces as potential sites for interracial interaction: community gardens in urban areas *Journal or Leisure Research* **36**: 36-355.

Shores, K. and West, S. (2010) Rural and urban park visits and park-based physical activity *Preventive Medicine* **50**: S13-S17.

Silva, T., Vaz de Melo, P. and Almeida, J. (2014) Revealing the city that we cannot see *ACM Transactions on Internet Technology* **14 (4)**: Article 26.

Silva, T., Vaz de Melo, P., Almeida, J., Salles, J. and Loureiro, A. (2013) A comparison of Foursquare and Instagram to the study of city dynamics and urban social behaviour, In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM.

Simmons, I. (2013) *Interpreting nature: Cultural constructions of the environment*. Routledge, London.

Smardon, R. (1988) Perception and aesthetics of the urban environment: review of the role of vegetation *Landscape and Urban Planning* **15 (1)**: 85-106.

Smith, L., Case, J., Smith, H., Harwell, L. and Summers, J. (2013). Relating ecosystem services to domains of human well-being: Foundation for a US index *Ecological Indicators* **28**: 79-90.

Söderström, O., Paasche, T. and Klauser, F. (2014) Smart cities as corporate storytelling *City* **18 (3)**: 307-320.

Soga, M. and Gaston, K. (2016) Extinction of experience: the loss of human-nature interactions *Frontiers in Ecology and the Environment* **14 (2**): 94-101.

Soga, M., Gaston, K., Koyanagi, T., Kurisu, K. and Hanaki, K. (2016) Urban residents' perceptions of neighbourhood nature: Does the extinction of experience matter? *Biological Conservation* **203**: 143-150.

Soga, M., Yamaura, Y., Aikoh, T. Shoji, Y., Kubo, T. and Gaston, K. (2015) Reducing the extinction of experience: Association between urban form and recreational use of public greenspace *Landscape and Urban Planning* **143**: 69–75.

Sotoudehnia, F. and Comber, L. (2011) Measuring perceived accessibility to urban green space: an integration of GIS and participatory map, In: *14th AGILE Conference on Geographic Information: Advancing Geoinformation Science for a Changing World*.

Srivastava, K. (2009) Urbanization and mental health *Industrial Psychiatry Journal* **18 (2)**: 75-76.

Stedman, R. (2003) Is it really just a social construction?: The contribution of the physical environment to sense of place *Society & Natural Resources* **16 (8)**: 671-685.

Su, S., Wan, C., Hu, Y. and Cai, Z. (2016) Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media *Applied Geography* **72**: 26-37.

Suh, B., Hong, L., Pirolli, P. and Chi, E. (2010) Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network, In: *Social computing, 2010 IEEE second international conference,* 177-184, IEEE.

Suminski, R., Poston, W., Market, P., Hyder, M. and Sara, P. (2008) Meteorological conditions are associated with physical activities performed in open-air settings *International Journal of Biometeorology* **52 (3)**: 189-197.

Sun, Y. (2017) Exploring potential of crowdsourced geographic information in studies of active travel and health: Strava data and cycling behaviour *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **42**.

Sun, Y., Fan, H., Helbich, M. and Zipf, A. (2013) Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation, In: *Progress in Location-Based Services*, 57-69. Springer, Berlin Heidelberg.

Sutko, D. and Silva, A. (2011) Combining web map services and opinion mining to generate sentiment maps of tourist locations, In: Jobst, M. (Ed) *Service Oriented Mapping 2012*. 277-286. JobstMedia Management Verlag Wien.

Swanton, K. (2008) *Healthy Weight, Healthy Lives: A toolkit for developing local strategies*. National Heart Forum / Cross Government Obesity Unit / Faculty of Public Health, United Kingdom.

Syrbe, R. and Walz, U. (2012) Spatial indicators for the assessment of ecosystem services: providing, benefiting and connecting areas and landscape metrics *Ecological Indicators* **21**: 80-88.

Tallis, H. and Polasky, S. (2009) Mapping and Valuing Ecosystem Services as an Approach for Conservation and Natural-Resource Management *Annals of the New York Academy of Sciences* **1162**: 265-283.

Tallis, H., Kareiva, P., Marvier, M. and Chang, A. (2008) An ecosystem services framework to support both practical conservation and economic development *Proceedings of the National Academy of Sciences of the United States of America* **105 (28)**: 9457–9464.

Talukdar, P. and Pereira, F. (2010) Experiments in graph-based semi-supervised learning methods for class-instance acquisition, In: *Proceedings of the 48th annual meeting of the association for computational linguistics*, 1473-1481. Uppsala, Sweden, Association for Computational Linguistics.

Tarman-Ramcheck, W. (1983) *Towards an ecosociology of ecocommunities* (PhD Thesis). Carroll University, Wisconsin.

Thelwall, M. (2014) Sentiment analysis and time series with Twitter, In: Weller, K., Bruns, A., Burgess, J, Mahrt, M and Puschmann, C. [eds] *Twitter and society,* 83-95. Peter Lang, New York.

Thoits, P. and Hewitt, L. (2001) Volunteer work and well-being *Journal of health and social behaviour* **42 (2)**: 115-131.

Thompson, J., Davis, S., Gittelsohn, J., Going, S., Becenti, A., Metcalfe, L., Stone, E., Harnack, L. and Ring, K. (2001) Patterns of physical activity among American Indian children: an assessment of barriers and support *Journal of Community Health* **26 (6)**: 423-445.

Thorsson, S., Lindqvist, M. and Lindqvist, S. (2004) Thermal bioclimatic conditions and patterns of behaviour in an urban park in Göteborg, Sweden *International Journal of Biometeorology* **48 (3)**: 149-156.

Thrift, N. and Pred, A. (1981) Time-geography: a new beginning *Progress in Human Geography* **5 (2)**: 277-286.

Tinsley, H., Tinsley, D. and Croskeys, C. (2002) Park usage, social milieu, and psychosocial benefits of park use reported by older urban park users from four ethnic groups *Leisure Sciences* **24**: 199–218.

Trepte, S. and Reinecke, L. (2011). *Privacy online: Perspectives on privacy and self-disclosure in the social web* Springer Science & Business Media, USA.

Troped, P., Saunders, R., Pate, R., Reininger, B., Ureda, J. and Thompson, S. (2001) Associations between self-reported and objective physical environmental factors and use of a community rail-trail *Preventive Medicine* **32 (2)**: 191-200.

Tsunetsugu, Y., Lee, J., Park, B., Tyrväinen, L., Kagawa, T. and Miyazaki, Y. (2013) Physiological and psychological effects of viewing urban forest landscapes assessed by multiple measurements *Landscape and Urban Planning* **113:** 90-93.

Tucker, P. and Gilliand, J. (2007) The effect of season and weather on physical activity: A systematic review *Public Health* **121 (12)**: 909-922.

Tudor-Locke, C., Bassett, Jr D., Swartz, A., Strath, S., Parr, B., Reis, J., Dubose, K. and Ainsworth, B. (2004) A preliminary study of one year of pedometer self-monitoring *Annals of Behavioral Medicine* **28 (3)**: 158-162.

Tumasjan, A., Sprenger, T., Sandner, P. and Welpe, I. (2010) Predicting elections with twitter: What 140 characters reveal about political sentiment *ICWSM* **10 (1)**: 178-185.

Turney, P. (2002) Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, In: *Proceedings of the 40th annual meeting on association for computational linguistics*, 417-424. Association for Computational Linguistics.

Twitter (2017) *Twitter usage/Company Facts*, https://about.twitter.com/company. Accessed 28.04.2017.

Tyrvainen, L., Ojala, A., Kalevi, K., Lanki, T., Tsunetsugu, Y. and Kagawa, T. (2013) The influence of urban green environments on stress relief meausres: A field experiment *Journal of Environmental Psychology* **38**: 1-9.

Tzoulas, K. and James, P. (2010) Peoples' use of and concerns about green space networks: a case study of Birchwood, Warrington New Town, UK *Urban Forestry and Urban Greening* **9 (2)**: 121-128.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kazmierczak, A., Niemela, J. and James, P. (2007) Promoting ecosystem health in urban areas using green infrastructure: A literature review *Landscape and Urban Planning* **81 (3):** 167-178.

Uitenbrook, D. (1993) Seasonal variation in leisure time physical activity *Medicine and Science in Sports and Exercise* **25 (6):** 755-760.

Ulrich, R. (1983) Aesthetic and affective response to natural environment, In: Altman, I and Wohlwill, J. [Eds] *Behavior and the natural environment*. 85-125. Springer, USA.

Ulrich, R. (1993) Biophilia, biophobia, and natural landscapes, In: Kellert, S. and Wilson, E. [eds] *The biophilia hypothesis*, 73-137. Island Press, Washington DC.

UN-HABITAT (2016) *World City Report 2016: Urbanisation and Development: Emerging Futures*. UN-Habitat, Nairobi.

Urry, J. (2007) *Mobilities*. Polity Press, Cambridge.

Van den Berg, A., Maas, J., Verheij, R. and Groenewegen, P. (2010) Green space as a buffer between stressful life events and health *Social Science and Medicine* **70**: 1203–1210.

Van der Spek, S., Van Schaick, J., De Bois, P. and De Haan, R. (2009) Sensing human activity: GPS tracking *Sensors* **9 (4)**: 3033-3055.

Vandenbulcke, G., Steenberghen, T. and Thomas, I. (2009) Mapping accessibility in Belgium: a tool for land-use and transport planning? *Journal of Transport Geography* **17 (1)**: 39-53.

Vicente, C., Freni, D., Bettini, C. and Jensen, C. (2011) Location-related privacy in geo-social networks *Internet Computing - IEEE* **15 (3)**: 20-27.

Wakamiya, S., Belouaer, L., Brosset, D., Lee, R., Kawai, Y., Sumiya, K. and Claramunt, C. (2015) Measuring crowd mood in city space through Twitter. In: Gensel, J. and Tomko, M. [Eds] *Web and Wireless Geographical Information Systems*, 37-49. Grenoble.

Walmsley, D. (1974) Positivism and phenomenology in human geography *The Canadian Geographer/Le Géographe canadien* **18 (2)**: 95-107.

Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S. (2012) A system for real-time twitter sentiment analysis of 2012 us presidential election cycle, In: *Proceedings of the ACL 2012 System Demonstrations*, 115-120. Association for Computational Linguistics.

Warburton, D., Nicol, C. and Bredin, S. (2006) Health benefits of physical activity: the evidence *Canadian Medical Association Journal* **174**: 801-809.

Warriner, A., Kuperman, V. and Brysbaert, M. (2013) Norms of valence, arousal and dominance for 13,915 English lemmas *Behavioural Research* **45**: 1191-1207.

Wells, N. and Evans, G. (2003) Nearby nature: a buffer of life stress among rural children *Environment and Behaviour* **35**: 311-336.

Wells, N., Evans, G. and Cheek, K. (2016) Environmental psychology, In: Frumkin, F. [Ed] *Environmental Health: From Global to Local*, 203-230. John Wiley and Sons, New Jersey.

Weng, J., Lim, E., Jiang, J. and He, Q. (2010) Twitterrank: finding topic-sensitive influential twitterers, In: *Proceedings of the third ACM international conference on Web search and data mining*, 261-270, ACM.

West, P. (1989) Urban region parks and black minorities: subculture, marginality, and interracial relations in park use in the Detroit metropolitan area *Leisure Sciences* **11 (1)**: 11-28.

Westphal, L. (2003) Urban greening and social benefits: a study of empowerment outcomes *Journal of Arboroculture* **29**: 137-147.

Whitehead, M. and Dahlgren, G. (1991) What can be done about inequalities in health? *The Lancet* **338 (8774)**: 1059-1063.

Wilson, E. (1984) *Biophilia*, Harvard University Press, Massachusetts.

Wolch, J., Byrne, J. and Newell, J. (2014) Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough' *Landscape and Urban Planning* **125**: 234-244.

World Health Organisation (WHO) (2017) *Urban green spaces: A brief for action*. WHO Regional Office for Europe, Denmark.

Wu, J. (2008) Making the case for landscape ecology an effective approach to urban sustainability *Landscape Journal* **27** (1): 41-50.

Wu, J. (2014) Urban ecology and sustainability: The state-of-the-science and future directions *Landscape and Urban Planning* **125**: 209-221.

Xia, C., Guha, S. and Muthukrishnan, S. (2016) Targeting algorithms for online social advertising markets, In: *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference.* IEEE.

Yardi, S. and Boyd, D. (2010) Tweeting from the Town Square: Measuring Geographic Local Networks, In: *Proceedings of the International Conference on Weblogs and Social Media*, AAAI.

Yli-Pelkonen, V. and Niemelä, J. (2005) Linking ecological and social systems in cities: urban planning in Finland as a case *Biodiversity and conservation* **14 (8)**: 1947-1967.

Zhan, X., Ukkusuri, S. and Zhul, F. (2014) Inferring urban land use using large-scale social media check-in data *Networks and Spatial Economics* **14 (3-4)**: 647-667.

Zhang, L., Riddhiman, G., Dekhil, M., Hsu, M. and Liu, B. (2011) *Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis*, HP Laboratories, HPL-2011-89.

Zhao, D. and Rosson, M. (2009) How and why people twitter: the role that micro-blogging plays in informal communication at work *Proceedings of the ACM 2009 international conference on Supporting group work,* 243-252. ACM.

Zickuhr, K. and Madden, M. (2012) *Older adults and internet use.* Pew Internet & American Life Project 6, United States.

Zimring, C., Joseph, A., Nicoll, G. and Tsepas, S. (2005) Influences of building design and site design on physical activity: research and intervention opportunities *American Journal of Preventive Medicine* **28 (2)**: 186-193.

Zuur, A., Ieno, N. and Elphick, C. (2010) A protocol for data exploration to avoid common statistical problems *Methods in Ecology and Evolution* **1:** 3-14.

Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., van den Hazel, P. and Brunekreef, B. (2010) Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route *Environmental Health Perspectives* **118 (6)**: 783.

Appendix 1. Example code used to create the Tweet corpus utilised in Chapters 3, 4, 5 and 6

As detailed in Chapter 2, the tweet corpus was gathered from tweets made publically available by Twitter through its REST API. R was used as the interface through which a connection to the Twitter API was made. The 'twitteR' package was used by the author as this has the necessary coding functions for downloading tweets already in place. The 'plyr' package was used in this process.

To make a connection through R possible, it is first necessary to create an account with Twitter. Using these account details an application account must be created with Twitter which can be done at the following website: https://dev.twissoutter.com/. This process generates a unique API key, API secret, access token and access token secret, which are needed to make a connection to the Twitter API. Once these are produced the following code can then be used to connect to the Twitter API using the oauth protocol.

```
> require(plyr)
> require(twitteR)
> consumer_key <- "……"
> comsumer_secret <- "……"
> setup_twitter_oauth(consumer_key, consumer_secret)
```

After a connection to the Twitter API is successfully made, the searching and downloading of tweets can begin. The following code provides an example of the process through which tweets were downloaded and stored prior to pre-processing. This example details the process through which tweets containing any mention of Cofton Park, one of the parks included in the sample of urban green spaces in this thesis, were searched for and downloaded.

```
> cofton <- searchTwitter("cofton park", n=1000, lang="en")
> coftonnrt <- strip_retweets(cofton)
> cofton.df <- twListToDF(coftonnrt)
> write.csv(cofton.df, file="Cofton Park.csv")
```

Appendix 2. Description of the manual annotation workflow used to assign emotion to tweets

This appendix provides the information given to the five manual annotators who were used to assign 1000 tweets into various emotion categories.

Annotation Manual

Please annotate each Tweet in the spreadsheet with one of the following emotions:

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise
- Beauty

If you cannot detect any of these emotions in the Tweet, please annotate it as *None.*

If the Tweet is not written at least partly in English, please annotate it as *not English*.

Annotation preparation

For this task, we define emotions as fitting categorical points of view. Though this may disagree with your intuition, it is necessary. Take a look at the figure provided and make yourself familiar with the categories. You may find this figure useful as a reference for you – If you find it difficult to pinpoint the appropriate category consult the more detailed emotional sub-categories.

Annotation Task

1. Read each Tweet carefully.
2. Try to identify the writer's emotion at the time of writing.
3. Assign the most appropriate emotion category by writing your choice in the emotion column to the right of the Tweet text.

    Note: you can use the text along with emoticons, hashtags and punctuation to identify the emotion. Ignore the URL links

    Note: if there are several emotions present, please select the strongest one
4. Once you have written your chosen emotion please move onto the next Tweet.

There are 1000 Tweets for you to annotate in total. You do not have to do them all at once, but please annotate them independently and do not discuss your choices with the other annotators until all 1000 have been completed.

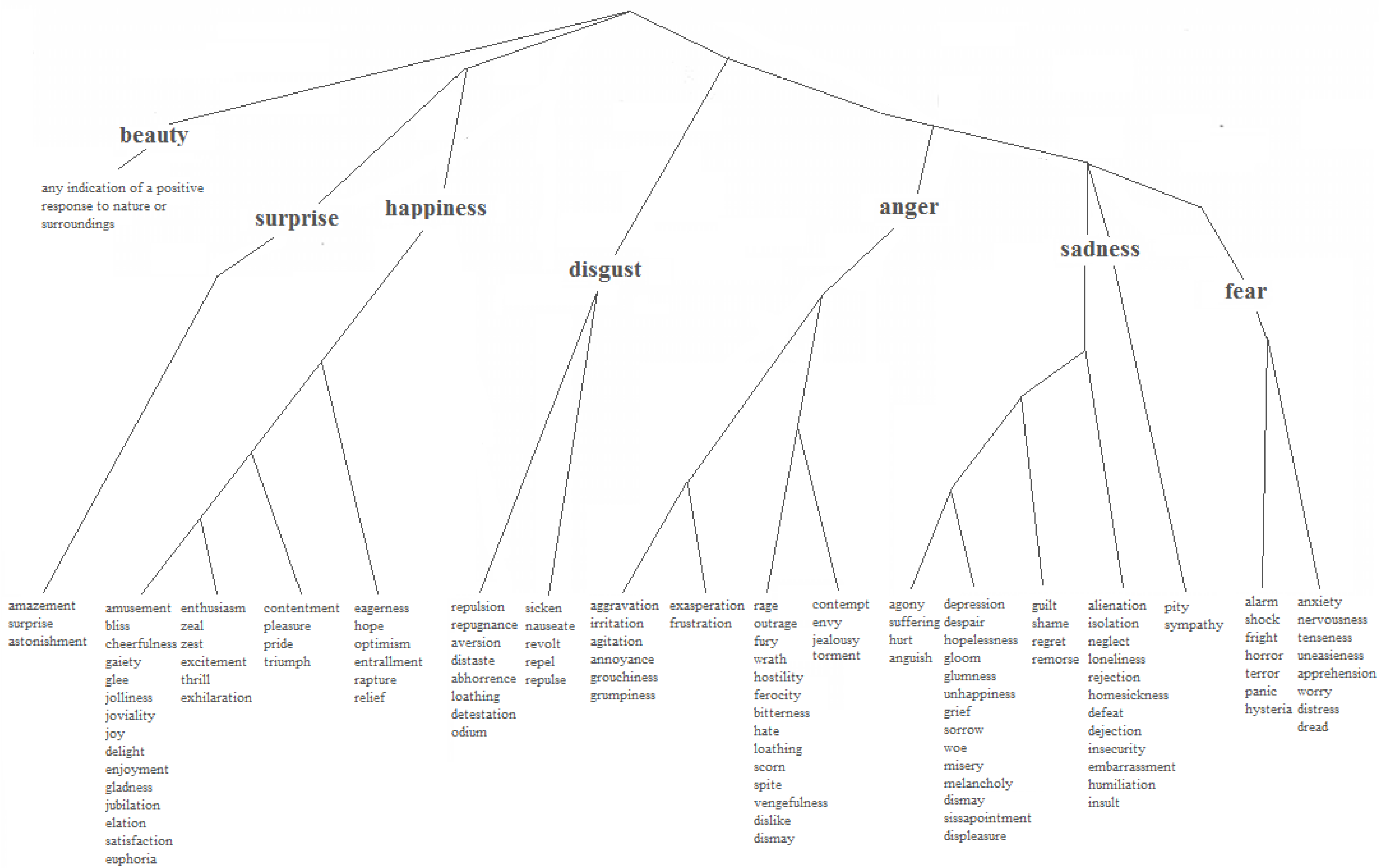


Figure 1: Use this as a reference to look up which category the emotion you have detected belongs to.

Appendix 3. Description of the semi-automated method workflow used to assign emotion to tweets

The semi-automated emotion classification method used in this thesis is an enhanced version of the Twitter Emotion Labeller (TwEmLab) developed by Resch et al. (2016) as part of the *Urban Emotions* project. It has since been refined to process large amount of social media data while producing similar emotion classification results. This appendix outlines the workflow and basic principles behind the algorithm. The core part of the classification workflow constitutes a graph-based semi-supervised learning approach in which a subset of manually labelled tweets (gold standard) are used as seeds to classify a large number of unlabelled tweets in a transductive learning process (Summa et al. 2016).

Pre-processing

Pre-processing removes all URLs and @usernames from the tweets. Any tweets found to be devoid of text at this stage are completely removed from the dataset. Additionally language detection is carried out and tweets which do not contain English words are also removed. A part-of-speech (POS) tagger and lemmatisation are then applied to the remaining tweets to prepare them for the subsequent linguistic similarity computation. POS marks every word in the text corpus with a corresponding POS tag (for example noun, common noun, adverb, verb, adjective). The Stanford CoreNLP NLP annotation framework is used for lemmatisation, which reduces the inflectional form of a word to a single term: the lemma. For example, lemmatisation transforms the example sentence “the girl’s ponies are different colours” to “the girl pony be differ colour”. Such linguistic reduction aids similarity computation for processing steps involving direct word comparisons.

Gold standard creation

The gold standard is a subset of the pre-processed dataset used to train the semi-supervised learning (SSL) algorithm. Five human annotators are used to annotate tweets into six basic emotion classes (*happiness*, *anger*, *disgust*, *sadness*, *beauty* and *fear*). *None* can also be used to annotate a tweet containing no or ambiguous emotion. Annotators are first required to label the same 261 randomly selected tweets and the results of this annotation are used assess reliability between annotators. Providing an annotator's Pairwise Kappa Index represents broad agreement (above 0.68, Fleiss, 1971), they are able to continue to the main annotation phase in which each annotator individually labels an additional 385 randomly selected tweets. Both annotated sets are merged to create the final gold standard. The gold standard is then split into a training (seeds) and test dataset.

Similarity computation and graph construction

In order to classify tweets according to any emotion present using SSL, a graph has to be constructed that encodes similarity between its nodes. In this context, similarity defines the likelihood of two tweets containing the same emotion. The seeds generated previously are used alongside the unlabelled tweets remaining in the dataset to construct an undirected similarity graph based on the textual dimensions of the tweets. TwEmLab assumes that basic emotion cases can be identified by using the linguistic characteristics of a tweet, thus similar linguistic characteristics encode similar emotion. Words, hashtags, unicode emojis, POS tags, spelling, punctuation and ANEW values are all used as linguistic characteristics from which similarity is computed. Using these linguistic features which characterise the words or style of a tweet, a similarity score is generated for every pairing of tweets in the dataset. Similarity is computed between every possible pair of tweets using a nested loop layout. The similarity scores calculated for each tweet combination are combined to overall similarity values, with each.

Graph-based semi-supervised learning and modified adsorption (MAD)

Once the graph has been constructed, an algorithm labels each node of the similarity graph based on a small amount of annotated seeds. MAD is the preferred algorithm in TwEmLab due to its increased effectiveness and scalability (Talukdar and Crammer, 2009; Talukdar and Cohen, 2014). It is an example of transductive learning in which there is no explicit training and label phase, and instead all instances are labelled at once. Graph based SSL is performed starting at the seed nodes and labelling each node of the graph in a random-walk process.

Evaluation

Finally, TwEmLab evaluates the results of the classification using a confusion matrix, precision recall and f-score values for all classes. Random and majority baselines are generated by assigning emotions randomly to all tweets or the most frequent class label to all tweets respectively. McNemar’s tests are used to identify significant differences between the classification results and these baselines.