Exploring the World with Volunteered Geographic Information: Space, Place and People

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Xining Yang

Graduate Program in Geography

The Ohio State University

2015

Dissertation Committee:

Daniel Z. Sui, Advisor

Ola Ahlqvist

Ningchuan Xiao

Abstract

We are undergoing a geospatial revolution with the rapid advances in technologies such as web 2.0, GIS, GPS, smart phones, and cloud computing, etc. The phenomenon of volunteered geographic information (VGI) is part of this revolution and has changed the traditional ways of how geographic data, information, and knowledge are produced and circulated. This dissertation presents a synoptic overview on how VGI, especially data from location-based social media, could be used as new data source for geographic research to better understand our world. Specifically, we argue that the emergence of VGI in geographic information science and technologies are revolutionizing basic and applied science by allowing integrated holistic approaches to reassessing the domain of GIScience by linking three themes together – space, place and people. Three interrelated case studies covering different conceptual linkages of VGI are presented as the urban dynamic in space and place, people's food choice and their exposed environment, and children beggars in China. We believe the ubiquity of volunteered geographic information has the potential to fundamentally widen the angle for geographers to better understand our world. In the coming data-driven era, we need to develop new conceptual framework, methodology and application to synthesize VGI with other geographic information.

Acknowledgements

Completing my Ph.D. degree is the most wonderful journey so far in my life full of challenges, unexpectedness and sweetness. I would like to extend my sincerest gratitude to the following persons who have offered guidance during the completion of this journey.

My first and foremost debt of gratitude goes to my advisor, Dr. Daniel Sui, for his persistent support in the past five years. I am so fortunate to have Dan as my advisor. His patient guidance, continuous encouragement, and invaluable advice make me finally achieve the academic goal. From the bottom of my heart, Dan serves as an inspiring figure and leads my way to this new realm of geography. His intellectual inspiration will have life-long influence on my academic career. His incredible time and energy devotion to research, teaching and students set a life-long high standard for me to follow.

My sincere appreciation also goes to my dissertation committee, Dr. Ola Ahlqvist and Dr. Ningchuan Xiao for their time and effort in refining my research design and improving my manuscript. I could not finish my dissertation without their informative input and careful feedback. I would like to say thanks to Dr. Catherine Calder for serving as my candidacy exam committee as well.

Meanwhile, I would like to thank my master's advisor, Dr. Yichun Xie; my bachelor's advisors, Dr. Yangge Tian and Dr. Fuling Bian; and my external mentor, Dr. Mike Batty for their kindness and willingness to support my career development. I also wish to thank my collaborators Dr. Xiang Chen, Bo Zhao and Dr. Xinyue Ye for their enthusiastic effort to work with me on research project. I learn a great deal through the collaboration.

Finally, my deep gratitude goes to my parents, for providing me with the opportunity to pursue my dream and for their unconditional love and support.

Vita

2004..Guangzhou No.2 Middle School, Guangdong, China

2008..B. Eng. Spatial Information and Digitized Technology, Wuhan University, China

2010..M.S. Geographic Information Systems, Eastern Michigan University

2010 to present ....Graduate Associate, Department of Geography, The Ohio State University, USA

Publications

Yang, X., Ye, X., & D.Z. Sui. (2015). We Know Where You Are—In Space And Place: Enriching the Geographical Context through Social Media. *International Journal of Applied Geospatial Research* (forthcoming)

Chen, X., & **Yang, X.** (2014). Does food environment influence food choices? A geographical analysis through "tweets". *Applied Geography*, *51*, 82-89.

Xie, Y., & **Yang, X.** (2011). Agent- Based Urban Modeling: Simulating Urban Growth and Subsequent Landscape Change in Suzhou, China. In Yang, X. (Ed.) *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, 347-357.

Fields of Study

Major Field: Geography

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Problem Context

Recently a new source of geographic information has become available through user-generated efforts, enabled by technologies loosely regarded as Web 2.0 (Goodchild, 2007; Vossen and Hagemann, 2010). Geographic information can now be found in the contents of wikis, blogs, tweets, photos, and many other forms of user-generated content. Goodchild (2007) defines this new source of data as volunteered geographic information (VGI), which emphasizes the role of voluntary efforts in producing geographic information, to differentiate with the conventional geographic information production process. Sui (2008) further points out that VGI has transformed the conventional way in how geographic information is created and used so it represents a “wikification” of GIS, which is not simply confined to new ways of data productions but also includes the development of GIS software, hardware, organizations, and people.

VGI and such citizen-driven data collection efforts are not new to Geography community (Elwood et al., 2012). Some historical related movements in Geography give rise to VGI. In the early 1930s, Stamp (1931) describes a group of volunteered teachers and students participated in the land use surveys in Britain. In the States, Bunge (1971) mentions in his “Geographical Expedition” that city residents were involved in a local counter mapping efforts. In the early 1990s, with the rapid development of geographic information systems, public participatory GIS (PPGIS) derives as a means to help researchers investigate ways of integrating and representing local knowledge of marginalized groups (Lin, 2013). In the Web 2.0 era, the active use of the Internet to promote citizen mapping has attracted huge attention from amateurs outside geography community. New forms of mapping and data collecting through public platform such as Google Map, OpenStreetMap, Arcgis.com generate tons of geographic information.

Compared to the traditional geographic data collected by government, federal agencies, or private sectors – e.g., U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) to represent roads, rivers as well as legal and statistical geographic areas, cell-phone company's call data to pinpoint a user to particular cell tower, and search engine query logs data with an associated IP address that can be inferred an approximate location of internet user – these volunteered geographic information are inherently self-reporting and public. As a result, they provide a rich and growing body of geo-tagged evidence that can potentially support scientific inquiries related to questions that geographers were difficult to deal with before. Such difficulties were often due to the proprietary nature of traditional location data, the cost of acquiring new data through small lab-based studies (e.g., due to navigating university IRB protocols, overcoming resistance to personal tracking devices), and the difficulties of sharing such sensitive data with other researchers.

Already we have witnessed the popularity of VGI has provided many research opportunities for geographers and GIScientists in the digital age. Many studies find out VGI to be a valuable source because of its huge potential to engage citizens and to be a significant, timely, and cost-effective source for geographers' understanding of the earth (Goodchild, 2007). People have witnessed the successful cases using VGI to help disaster relief by providing useful location-based information to the victims after the 7.0 earthquake in Haiti in 2010 and the Great East Japan Earthquake and Tsunami in 2011 (Zook et al., 2010; Yap et al., 2012).

One major drawback of current GIS system is the underrepresentation of Place (Goodchild, 2011) and People (Miller, 2005). Space and place are two fundamental concepts in geography, and more broadly in social sciences, humanities, and information science. Space is more abstract and generic while the notion of place is more tangible to humans. Efforts of VGI also have powerful implications for studies of a fundamental geographic concept *place*, since information contributed from the average citizen can potentially help us to redefine and thus formalize associated concepts of space and place. Until recently, Geographic information science has been dominated by perspectives from space using Cartesian coordinates according to Euclidean geometry and it has often been criticized for promoting an absolute and mechanistic representation of space that has limited its usefulness for representation and analyses of human-centered issue (Pickles, 1995). Goodchild (2011) also points out that GIS are typically space-based and are not effective in representing the "sense of places". The massive amounts of volunteered geographic information (VGI) in general, and geo-tagged or location-based social media data in particular, seem to revive our approach to the world from the perspective of place (Haklay, 2010). By linking VGI research to the traditional dichotomy in geography between space and place (Tuan 1977), Elwood et al. (2012) indicate that VGI research invites a more place-centric perspective, in contrast to much previous concentration in geographic information science (GIScience) with space. Moreover, as VGI is created by citizen volunteers, it tends to favor the character of human discourse rather than the scientific measurement, thus have strong links to the world of place and people (Elwood et al., 2012). The recent convergence of GIS and social media (Sui and Goodchild, 2011) further prompts a new level of urgency for theoretical works to reconcile the world of space (traditional GIS) and the world of place (social media). How to formalize place in the GIS context will be both interesting and challenging. Thus this project undertakes the challenge by shifting our attention in GIScience from space to place and people.

1.2 Research Objectives

VGI has continued its booming trajectory during the past 5 years as a result of the development of Web 2.0 and location-based services. Especially, the world have continued to witness that the convergence of GIS and social media has become more mutually constituted with online mapping sites becoming increasingly social and social networking sites more location-based (Sui and Goodchild, 2011). The huge increases in user-generated content (UGC) enabled by Web 2.0 offer many new possibilities for analysis, and researchers are beginning to exploit these burgeoning online data sources (Miller 2010; King 2011) and how to dig in with these “big data” (Crampton et al., 2013). While there exist some existing research on the VGI and its use as an important data source in geographic studies, methods for using these data, particular from a place perspective, are not broadly discussed.

The objective of this study therefore shall be to contribute to existing theories and methods of developing a framework to utilize volunteered geographic information, geo-social media data in particular, as a new geographic data to better understand our world by linking space, place and people.

What kind of geographic information can be extracted from social media and how to harvest these VGI?

What are the dynamics of human behavior and pulse of a city that we can capture by using geo-social media data?

How can VGI be used to understand the world from a place perspective? What are the correlation between the VGI and the socioeconomic characteristics from local residents and communities? What light do VGI shed on social issues such as individual health and societal inequality?

1.3 Synopsis of the Dissertation

The thesis is organized as follows:

Chapter 1 is the introduction of the study.

Chapter 2 will review the origin, type and current research of volunteered geographic information (VGI) as well as examine space and place from a geographical perspective. By doing this I will setup the conceptual framework of this dissertation.

Chapter 3 will outline the mechanism on how to harvest geo-social media data that is widely used in my dissertation along with a subsequent methods of exploring the urban dynamic, leading to a case study of the pulse of a city within the Columbus, Ohio MSA using social media checkin data.

Chapter 4 explores the applicability of geo-social media data on studying food accessibility from a people-based perspective. This chapter provides a novel method to use VGI as data source to study food-related issue on individual level and a 5-weekday period case study in Columbus, OH reveals some interesting finding on people's choice of food and their exposure to food environment.

Chapter 5 presents how VGI has been used to address a social inequity issue – street child begging and missing child, that has gained widespread attention on Chinese social media. Using a representative sample of microblog posts as well as contextual information from a variety of sources, this chapter examines the geographic distribution of street children in China as well as identifies how VGI are capable of empowering citizens to help those children to find their way back home.

Chapter 6 is the conclusion of the dissertation. This chapter concludes this doctoral study, indicates its limitation, and provides potential boulevards for future study.

Chapter 2: Literature Review and Conceptual Framework

2.1 The Origin, Type and Current Research of VGI

2.1.1 The Origin of VGI

Volunteered Geographic Information (VGI) is defined as "*geographic information acquired and made available to others through the voluntary activity of individuals or groups, with the intent of providing information about the geographic world*" (Elwood et al., 2012 p. 575). The growth of VGI in the past 10 years is a result of several related technological advances and scientific practices, such as web 2.0, geoweb, spatial media, neogeography, citizen science, crowdsourcing, and open science (Rice et al., 2012). Goodchild (2007) coined the term VGI to describe what he viewed as a different form of spatial data production that contrasted with more conventionally produced geographic information, and to emphasize the potential role of VGI in augmenting our knowledge of the geographic world. VGI often differs from conventionally produced forms of geographic information, including the types of information produced and the approaches used to acquire it, the methods and techniques for working with it, and the social processes that mediate its creation.

Despite the increased attention to VGI in recent years, voluntary spatial data collection efforts are not new and can be traced back before the age of web 2.0. A host of related movements have helped enable VGI as a phenomenon. One of the earlier examples was the Christmas Bird Count starting back in 1900. In geography, Bill Bunge's 'Geographical Expeditions' involved local residents in counter-mapping and spatial data production efforts in Detroit in the 1970s (Bunge, 1971). In the U.K., the BBC's Domesday Project in the 1980s used volunteers and community groups to assemble a digital spatial data archive (Openshaw et al., 1986). The explosive growth of VGI in the past decade is a result of several converging technological advancements. The traditional high cost of spatial data collection – which necessitated its production either by government agencies or corporations – has now been vastly reduced due to broad access to the portable global positioning system (GPS) devices, location-aware smart phones and other handheld/wearable devices, broadband Internet, and the availability of user-friendly mapping software.

2.1.2 Types of VGI

There have been several efforts to develop VGI typologies according to the purpose of VGI (allocentric vs. egocentric) or the nature/content of VGI (implicit vs. explicit) (Craglia et al, 2012; Engler et al, 2014). However, in the age of big data, the boundary between VGI and other types of geographic information/spatial data is rapidly blurring. In fact, as a special type of geographic information, VGI shares multiple commonalities with the traditional geographic information we have been dealing with so far. In general, VGI can be loosely grouped into three types: *geospatial framework data, gazetteer/place name data, and miscellaneous geo-tagged content data (text, audio, photo, and video etc.) about various places/points of interests.* Furthermore, VGI can be used to augment existing authoritative geospatial databases of various kinds.

- Geospatial Framework Data: OpenStreetMap[1]

The production of transportation and road network data using the VGI approach has been successful. OpenStreetMap (OSM), perhaps the best example of geospatial framework data collected by volunteers, is a collaborative global effort to produce a detailed, full coverage digital map of the world's road networks (and some other features) through voluntary effort. Volunteers capture the locations of transportation infrastructure and topographic features by using GPS and tracing satellite imagery. Once compiled, these data are freely available as rendered online maps or for download as data files under the framework of creative commons. Users are free to use these data and to develop their own applications with it as long as they acknowledge their use of OSM data.

- Gazetteer/Place Name Data: Wikimapia[2]

Wikimapia harnesses efforts from volunteers to "describe the whole world" (the website mantra) by integrating a wiki system with Google Map services to enable users to identify and describe Earth-surface features. While gazetteer information is limited to name, feature type, and location, most entries on Wikimapia offer richer descriptions, and may provide links to external resources for more detailed information. As of 2011 nearly 8,000,000 features have been described, which is beyond the number of entries of the largest extant gazetteer. In recent decades, high costs have caused dwindling augmentation and maintenance of authoritative gazetteers, where Wikimapia offers a viable substitute.

---

[1] www.OpenStreetMap.org
[2] www.Wikimapia.org

- Geo-tagged Content Data: Google Map[3]

Google Map is a web mapping service application and technology provided by Google providing street maps, satellite imagery and street viewer functions. Besides its fundamental mapping role such as place looking and route planning, perhaps even more significantly, Google's new mantra "Google Maps = Google in Maps" has made all geo-tagged on-line VGI content searchable through Google maps, including photos, texts, videos, music etc., which is revolutionary not only for geographic data production, but also for the design and development of search engines. More than just an easy way to georeference photos, this is potentially a valuable source of VGI.

2.1.3 Current Research of VGI

Since the official debut of the term "volunteered geographic information (VGI)" in the literature (Goodchild, 2007), there have been meetings, workshops, special issues on academic journals devoted exclusively to this theme. Examples include the 2007 NCGIA VGI workshop, the AutoCarto 2008 workshop, the USGS 2010 VGI workshop, the GIScience 2010 VGI workshop, the 2011 VGI Pre-Conference at AAG, the GIScience 2012 VGI workshop, *GeoJournal* (Elwood, 2008a, b), *Journal of Location-Based Services special VGI issue* (Rana and Joliveau, 2009) and *Geomatica* (Feick and Roche, 2010). Three pioneering researchers in this field have coedited the book titled *Crowdsourcing Geographic Knowledge* (Sui et al., 2013) taking stock of recent advances in VGI research. Based on different perspectives of the studies, recent research about VGI could be generally categorized into three following groups.

---

[3] maps.Google.com

The first group of research care about the VGI data themselves. The objective of their study is to take an inventory of VGI production and conduct quality assessment for the major types of VGI currently available online. For example, a pilot study (Figure 1) conducted by research group at UCSB collects around 100 websites of VGI initiatives in 2009. Then they assessed the intended geographic extent of the information collection effort, date the initiative was begun, type of organization or group initiating the project, and primary purpose of the initiative of these websites (Elwood et al., 2012).

VGI Sites Inventory

Link	Title	Description	Since	Geo-coverage	Who	All terms	Structure
view	Birds and Climate change: building an early warning system	citizen science data about birds indicating climate change	mid 2000s?	Regional	South African National Biodiversity Institute and University of Cape Town	report birds climate change citizen science	
view	Geocrowd	With Geocrowd we propose geoblogging as a means for spatiotemporal...	??	Local	??	geoblogging	
view	Carticipate	Carticipate is an experiment in social transportation, the first and only...	?	Local	?	GSN	
view	Google FluTrends	We've found that certain search terms are good indicators of flu activity...	2009?	Local	Google.org; USA; non-profit	data mining, GIM	
view	Fire Eagle	Fire Eagle is a site that stores information about your location. With your...	2009	Local	Yahoo; USA; for-profit	GIL	
view	Yahoo! Placemaker	Placemaker provides geo-enrichment for the hugely significant proportion of Web...	2009	Global	Yahoo; USA; for-profit	GIC	
view	Glympse	Glympse visually answers the question of "Where are you?" with a...	2009	Local	Glympse; USA: for-profit	GIL	
view	Wild Style City	a virtual graffiti application that turns the urban environment into a canvas...	2009	Local	earthmine, inc.; USA; for-profit	GSE, virtual graffiti	

Figure 1 VGI Sites Inventory[4]

[4] http://vgi.spatial.ucsb.edu/inventory

This study offers a systematic overview of the new phenomenon rather than an up-to-date comprehensive summary of VGI websites, since voluminous VGI projects have been conducted or ceased since this 2009 inventory. There are also some studies discussing how to ensure the quality of VGI, since it's a voluntary effort. Goodchild (2009) argues that the richness of geographic context makes it difficult to falsify VGI thus it is relatively easy and accurate to place a volunteered fact about some location into the context of existing information. Another study finds out that despite VGI doesn't typically include traditional measures of accuracy or inaccuracy and it lacks of metadata, research conducted by Haklay (2010) contests the accuracies of OpenStreetMap are comparable to those of authoritative sources. To further assure the quality of VGI, a recent study by Li and Goodchild (2012) summarizes three different approaches (i.e., crowd-sourcing, social, and geographic) to quality assurance.

The second group of recent studies focuses on developing methods for using VGI. As is mentioned in the first part, the explosive growth of VGI is proven to be a valuable data source while the large volume and complexity of VGI have made it almost impossible to analyze or synthesize manually. Productive use of these new data source with spatial and temporal attribute relies on developing new analytical techniques in the context of spatial data mining and geographic knowledge discovery (Alvarez et al., 2008). Many VGI applications in the context Web 2.0 come in the form as map mashups (Batty et al., 2010), a term borrowed from the music industry originally referred to a composition created by blending two or more songs. Such activities of using VGI are greatly reported in disaster reliefs globally. For example, Liu and Palen (2010) illustrate a qualitative research of the design and creation of crisis-map mashups to describe emergent neo-geographic practices in emergency management and disaster relief. They first analyze the situations which include mashup creation, data selection, and design of spatial and temporal information representation. Moreover, they further discussed the implications of emergent neogeographic practices using Ushahidi and New Orleans repopulation maps as case studies to illustrate benefits obtained by synthesizing professional, paleo participatory and neogeotechnologies for crisis mapping. Lin (2013) also provides some VGI initiatives in China and among those one is how a Google Map mashup map is created and help the relief support and needs in the Sichuan earthquakes happened on May 12, 2008. This VGI application received 82,539 hits within a week after its creation and it soon reached a million hits and it was used by some NGOs on the site of disaster to assist their relief efforts. Furthermore, VGI and user-generated content (UGC) are increasingly becoming data source in addition to the traditional sources for geographic research. Cidell (2010) reports the use of content clouds (a way to summarize the contents of a document by depicting the words according to its high to low frequency in larger, darker type within the cloud) as a method of exploratory qualitative data analysis using primarily online information. When used as a form of qualitative GIS, content clouds provide a meaningful way to summarize and compare information from different places on a single issue. Using people-finder sites (such as Whitepages.com) and social networks (Facebook) as well as information contributed as volunteered geographic information (VGI), Chow (2013) outlines a research agenda to explore the web demographics using VGI. Using multiple online people-finder sites and VGI information, he proposes a framework to harvest, process and validate personal-level web demographics including personal details, marriage status, financial history, etc. Another example of using VGI is the Geo-Wiki project (Fritz et al., 2009) initiated by a group of physical geographers and earth scientists. Similar to the OpenStreetMap project, the Geo-Wiki project aims to gather volunteers all around the world to help improve the quality of global land cover maps. Such crowdsourcing effort attracts volunteers from multidiscipline and they collaboratively produce large amount of inexpensive ground-truth data and knowledge which is valuable for the purpose of validation and the future creation of global land cover maps.

Another track of research to VGI is to examine its social impact. Besides the roles as an important new data source, VGI is also conceived as a new social practice with attention to ways in which the processes, relationships, and products of VGI activities structure and represent knowledge and shape social and political relations (Elwood et al.,2012). By studying VGI from a different perspective, some social issues have surfaced. First of all, many of research have focused on the issue of digital divide, examining the inequalities among people and places by studying the origination and distribution of VGI. Crutcher and Zook's (2009) illustrate patterns of inclusion and exclusion in VGI which mirror the prior research of the digital divide by studying user-generated information in Google Maps. Their findings suggest a necessity to reconceptualize the mechanisms in the context of VGI. To further study the digital divide using VGI, Baginski et al. (2013) conduct an empirical study in Franklin County, Ohio. They retrieve one form of VGI—restaurant reviews—to evaluate the digital divide within the Columbus MSA and their finding shows that Web 2.0 impacts have not touched down evenly within the study area. Second, privacy and surveillance are significant concerns associated with the phenomenon of VGI as well. Geographic information contributed by users to a website or database enables the ability for location tracking and monitoring, which may lead to an invasion of privacy and the potential for harm (Andrejevic 2007). Many social media sites and location-based services record the locations of their users in space, which is sometimes actively provided by the user, however in other cases it is produced passively without the user fully understanding this or its consequences. Some location-based services such as Foursquare utilize this location information in developing their service, while others compile it for use by other companies for targeted geo-demographic marketing purposes (Boyd and Crawford 2012, Wilson 2012). Willingly or unwillingly sharing one's location – whether for the purposes of contributing VGI or not – is a serious safety concern, and presents one of the most serious issues implicated with the recent integration of mobile and location technologies (Elwood and Leszczynski 2011). In crowdsourcing and VGI projects, users should be made aware of what the data will be used for, and any potential privacy infringement related to their contributions. Last but not least, intellectual property is another potential legal concern for various VGI practice. According to Scassa (2013), these legal VGI issues may manifest in different ways depending on the stakeholders concerned. The operator or producer of a product, website, or application that relies upon VGI will need to be aware of various potential legal issues, and should take care to address these issues in licenses and other documentation related to the site. The extent to which they need to do so and how they do so may depend on whether they are public or private actors, or whether they operate commercially or non-commercially. VGI contributors should be aware of the license terms under which they provide content. Although they are unlikely to hold copyright in small quantities of information, creative material provided in other forms, most notably photographs, video, or text, is likely protected by copyright. In some cases, contributors of VGI may be institutions or organizations that provide data sets, and there may be a copyright that can be asserted in an original selection or arrangement of data.

2.2 Space and Place in the Geographic Perspective

Space and place together define the nature of geography (Tuan, 1977). The concept of space and place has a long history in geography and related discipline but has been plagued by a fundamental vagueness of definition (Goodchild, 2011). As Tuan (1977) and Casey (1997) have demonstrated, scholars in multiple fields throughout history have developed a vast repertoire of conceptualizations of space and place. Conceptually, space and place represent a continuum of how we understand the world and my understand is that space is a more abstract concept and place is more concrete. There are three different conceptions of space that have been drawn on in geography – absolute, relative and relational. According to Agnew (2011), the concept of place usually includes three pillars: location (as defined latitude and longitude), locale (as defined by both physical/environmental and socioeconomic/cultural context), and a sense of place (as defined by human subjective perception/attachment to a particular location/locale).

Apparently, place is semantically a much richer concept than space. In GIScience, space and place are sometimes used interchangeably while in fact; the distinction between space and place can be traced back to ancient Greeks who made a distinction between chora (measurable, abstract space) and topo (immeasurable, concrete place). According to Curry (2005), geography was understood for quite some time primarily from the perspectives of chora and topo. “Geo” entered into the scholarly and popular lexicon at a much later age in human history. Casey (1997) further observed that Western thinking about the world had continuously neglected the perspective of place and been increasingly in favor of space since Enlightment. Scholars now seem to have consensus that a partial focus on either space or place might serve as an impediment to a more holistic understanding of the world.

The space-place debates in geography started in the late 1960s from a spatial scientist John Hudson when he was asked what geographers did and he replied as “dots and lines”. This answer triggers the debates of geographers trying to revise Hudson’s answer in the following years. John Agnew (2005) categorizes three broad approaches to the space and place distinctions and debates. (1) place over space; (2) space over place; and (3) the very distinction is superseded by instead “putting space and place together'” (Agnew, 2005, page 89).

- Place over space

Tuan (1977) wrote in his paper “What begins as undifferentiated space, becomes place as we get to know it better and endow it with human value. ... Place is pause; each pause ... makes it possible for location to be transformed into place (page 6).” This represents a school of opinion that place is good because it can embrace humanistic value and sentiments thus turning location from a set of abstract coordinates into a site with meanings (Wainwright and Barnes, 2009). In this view space only provides the raw material for place as Relph (1976) writes, “space provides the context for places” (page 8). Under this view space is only derive its meaning from particular places and thus place dominate space.

- Space over place

Space trumps place in this view. David Harvey (1985) studies the creation of distinct spaces of capitalist economic circulation and accumulation and he argues that space was where the action occurred and places were only staging posts for spatial transformation. Later Harvey (1996) further develops his scheme and point out that change in space is constant, primary and resurgent which follows space’s involvement in capital mobility while place is in a secondary role and is only the effects of the spatiality of capital circulation within which they are produced. Needs a bit more elaboration.

- Relational space and place

In this view space and place are linked together using a relational approach. Doreen Massey (1991) and Marcus Doel (1999) want to wipe out the space-place divide and they even invoke a new word “splace”. Massey (1991) argues that space and place are relational and they are the consequence of the connectedness of things. She believes that places are made from the same kinds of relations that produce space and she denies any essential difference between them. Similar effort by Doel (1999) proposes that we should abandon the space-place distinction.

My opinion on space and place might be situated on Agnew (2005) observation that space could be considered as 'top-down,' defined by powerful actors imposing their control and stories on others. Geography practiced from the space tradition tends to be a vertical perspective, quantitative, and predominantly visual (appealing to the eye). In contrast, place can be considered as 'bottom-up,' representing the outlooks and actions of more typical folks (Sui, 2011). Geography practiced from the place tradition favors a slant/side perspective (e.g. Google Streetview or regular photos), qualitative, and full of context. Wainwright and Barnes (2009) provide some common terms used to distinguish space and place (table 1).

Table 1 The space-place distinction

Space	Place
Motion	Pause
Extension	Community
Position	History
Totality	Particularity
Infinity	Identity
Exteriority	Interiority

2.3 Proposed Framework linking VGI to Space, Place and People

Goodchild et al. (1999) propose a theoretical framework to understand the three distinct arenas that GISciences should address (Figure 2).

- Cognitive Models of Geographical Space (issues of the individual): how do people conceptualize the world around them, and reason about it using those conceptualizations?
- Computational Methods for Representing Geographical Concepts (issues of the system): how can we design GIS to achieve maximum performance and functionality, with minimum information loss or other constraint?
- Geographies of the Information Society (issues of society): what processes determine the adoption of GIS in society, and its use by institutions, and how does the adoption of GIS change the way society constructs space?

The emergence of VGI in Geographic information science and technologies are revolutionizing basic and applied science by allowing integrated holistic approaches to reassessing the domain of GIScience, as proposed fifteen years ago, with new perspectives. The theoretical framework this dissertation proposes as a base upon which to develop and apply quantitative and qualitative geographic method that enable understanding the world with volunteered geographic information draws upon a wide range of perspectives from human geography and social sciences more generally. The three pillars focused on here are *space*, *place* and *people*, each of which has been addressed in the past with specific methods and perspectives within GIScience community (Tobler 1970, Goodchild 2011, Miller 2005). All three themes proposed here demonstrate strong conceptual overlaps and stem from the previous literature (Goodchild et al., 1999) on examining components of geographic information science.

Figure 3 demonstrates the framework of volunteered geographic information with linkage to *space, place and people*. The inner ring denotes the variety of VGI as I elaborate in section 2.1.2. This represents a new kind of geographic information for scientific research and application. This intersects with three outer rings, which correspond to three conceptual themes – *space, place and people*. The intersection of VGI with each theme is further elaborated as potential data source and application in each rectangle.

- VGI and Space

Most spatial analysis is based on an explicit or implicit interpretation of the First Law of Geography (Tobler 1970, P. 236): "Everything is related to everything else, but near things are more related than distant things." Essentially as a geospatial data with geo-tag, spatial patterns could be identified from various volunteered geographic information and numerous models were constructed to interpret the observed patterns (Li et al., 2013; Zook and Poorthuis, 2014).

- VGI and Place

The availability of VGI from such applications may have significant implications in the future of qualitative GIS research as a growing body of work has become interested in drawing upon alternative methodologies to meet the limits of conventional GIS in which phenomena become decontextualized place. In other words, the rich content produced through location-based social media offers the possibility to facilitate geographic research that illuminates the importance of place perceptions (Li and Goodchild, 2012; Purves et al., 2011).

- VGI and People

The emergence of VGI also provides us a golden opportunity to develop what Miller (2007) envisioned as people-based geographic information science. Conventional GIScience technology such as remote sensing image classification has been successfully applied in revealing physical geographical features such as land cover information and climate change. On the contrary, VGI and the current geospatial big data in general well capture human activities. For example, large volumes of spatio-temporally tagged volunteered geographic information lead to the upsurge of human mobility research (Lu and Liu 2012; Yue et al. 2014). Another track of research area is to study human's health in geographic context (Yang and Mu, 2014; Mitchell et al., 2013).

The linkages between the two rings suggest that using VGI as a new geographic information may provide new insight into human geography and social science, thus further nurturing three corridors with theoretical guideline for chapter 3, chapter 4 and chapter 5 respectively. In the following three chapters, I construct three data-driven empirical study to reflect each of the proposed VGI research corridor respectively.

- VGI, Space and Place: Use VGI to Study Urban Dynamics in Space and Place
- VGI, Place and People: Use VGI to Study Individual Food Environment
- VGI, Space and People: Use VGI to Study Child Beggars in China

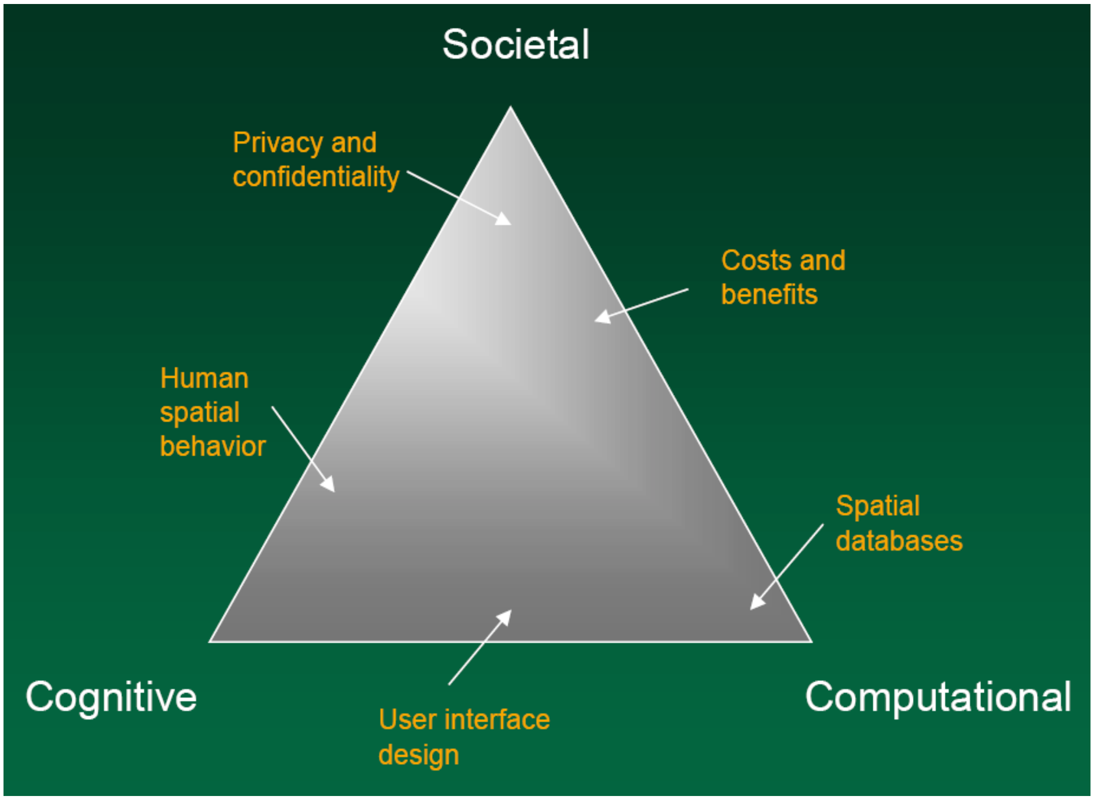


Figure 2 Geographic Information Science Conceptual Framework

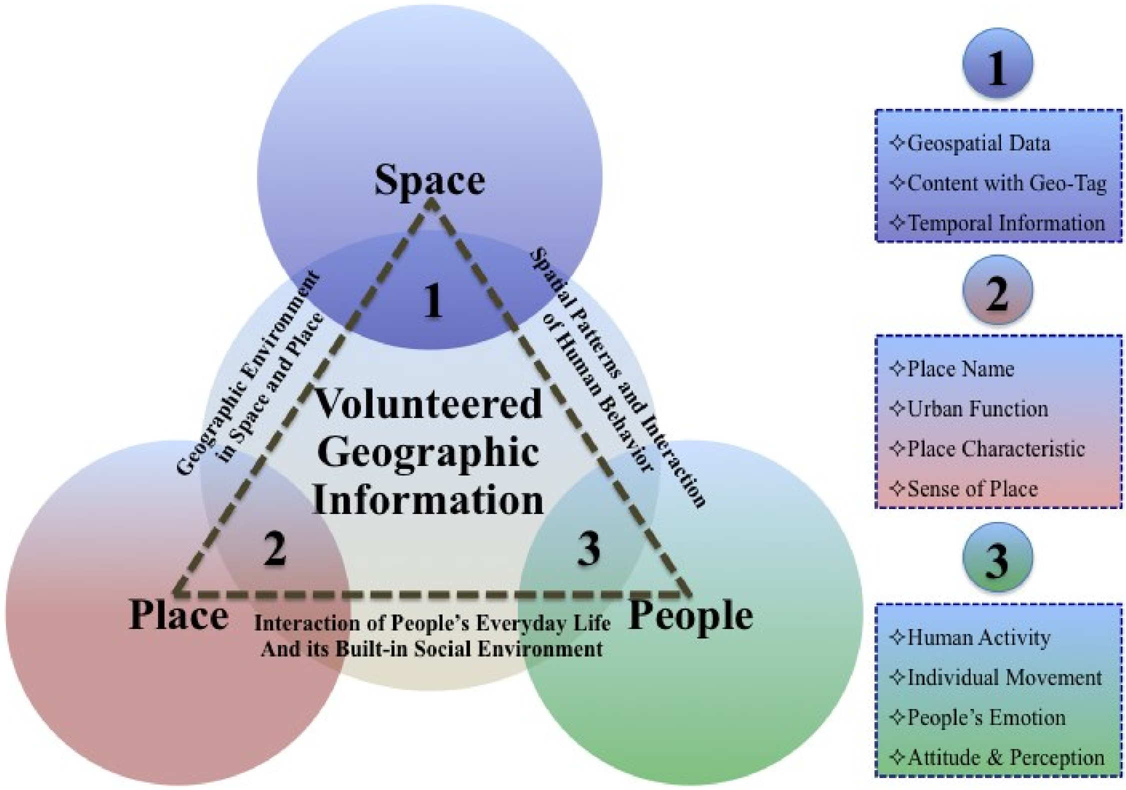


Figure 3 Proposed Framework to Explore the World with Volunteered Geographic Information

Chapter 3: Use VGI to Study Urban Dynamics in Space and Place[5]

Understanding urban dynamics is crucial for a number of sub domains in geography, but it can be expensive and time consuming to gather necessary data. The rapid rise of volunteered geographic information in general, and geo-social media data in particular, has given us a new massive source of geo-tagged data that can be transformative in terms of how we understand our cities. This chapter will first demonstrate the working flow on how to harvest geo-social media data that is widely used in my dissertation followed by presenting subsequent methods of exploring the urban dynamic in space and place, leading to a case study of Columbus, Ohio MSA using social media checkin data.

3.1 Introduction

Over half of the world's population now lives in cities (Martine and Marshall, 2007) thus majority of human and natural events in motion happen in cities. Understanding the cities we live in has never been more important. Until quite recently we relied upon human recording of these motions which inevitably incorporate the perceptions, fallibilities, and limits of ourselves as observers (Batty, 2010). Moreover, a number of methods that are used to collect data about urban dynamics and people activities tend to be slow, labor-intensive, expensive, and lead to relatively sparse data.

[5] Portions of this chapter will be published in Yang et al. (2015)

This began to change with the development of the Internet, Web 2.0 and Broadband communication technology. Its access through wireless devices with geography positioning systems (GPS) that can be incorporated into ever-smaller recording instruments such as handheld GPS device or smartphone is now enabling individuals to record their motion at specific locations and on a continuous basis. We argue that there is an exciting opportunity for creating new ways to conceptualize and visualize the dynamics, structure, and character of a city by analyzing the social media its residents already generate. Millions of people already use Twitter, Instagram, Foursquare, and other social media services to update their friends about where they are, communicate with friends and strangers, and record their actions. Some of this media is geo-tagged with GPS data, making it possible to start inferring people's behaviors over time. We believe that this kind of geo-tagged social media data, combined with new kinds of analytics tools, will let geographers, urban planners, policy makers, social scientists to explore human's activity as well as the pulse of a city, in a manner that is cheap, highly scalable, and insightful.

Space and place are two fundamental concepts in geography that together define geographers' view on the world (Tuan, 1977). Both concepts have a long history in geography and related disciplines. Scholars in multiple fields throughout the history have developed a voluminous literature on conceptualizations of space and place (Wainwright and Barnes, 2009; Agnew, 2011; Casey, 2013). Conceptually space is often related to an abstract view perceived as a "top-down" while place is related to be a more concrete one considered as a "bottom-up" (Agnew, 2005; Sui, 2011). Partial focus on either space or place might serve as an impediment to a more holistic understanding of the world from a geographic perspective (Sui, 2011). The difference between space and place has been mainly discussed by human geographers, but received little attention in GIScience as these two words have been used interchangeably in the GIScience literature. In fact, until recently the space perspective has dominated GIS and geographic objects are mainly represented using Euclidean geometry and Cartesian coordinate system (Sui, 2011). In contrast to space, place can be considered as a “bottom-up” view that focuses on capturing the local environment and human activity in a qualitative manner (Sui, 2011). Some pilot work has begun in shifting from space to place in GIS practices (Goodchild, 2011). Indeed, how to incorporate place into GIScience will be both interesting and challenging.

The convergence of social media and GIS (Sui & Goodchild, 2011) provides an opportunity to reconcile the world of space (traditional GIS) and the world of place (social media). Until recently all geographic information was produced mainly through a top-down process by government mapping agencies or corporations in the mapping industry. The emergence of social media such as Twitter, Facebook and Foursquare and the widespread adoption of GPS-enabled tagging of social media content provide new opportunities for us to study activities of people from a place perspective. These internet-based micro-blogging applications allow users to post and read short messages related to local environment through a shared social connection (Boyd and Ellison 2008). Such data usually present with fine spatial-temporal granularity and sometimes have rich context such as text, photos and other qualitative information. The geography of social media data such as the user or messaging location is invaluable for researchers to link where we are to what we do or feel, using real-time or archived data.

Spatially identifying user-generated conversations on social media may implement policy more effectively to where the targeted audiences live (Ghosh and Guha, 2013). Scellato et al. (2011) try to predict social link by detecting user patterns in Foursquare, a popular location-based social network. In addition, Cramer et al. (2011) aim to discern users' location sharing patterns from the individual user's perspective. As Jiang and Miao (2014) argue, "Location-based social media enable users to track individual historical trajectories, their friends, and even the growth of social media. Unlike with conventional cities, the trajectories of social media are well documented by the hosting companies; and unlike conventional census data, social media data is defined at individual level, often at very fine spatial and temporal scales" (page 2). The explosive growth of location-based social media in recent years provides us an unprecedented opportunity to explore human activities in both space and place.

In this chapter, social media data are crawled from a location-based service website Foursquare[6] via twitter public stream API. The goal of this chapter is to conduct an empirical study in the GIS environment aiming to enrich both the spatial and platial context of social media data, using checkins as an example. We explore three main research questions:

- What spatial and temporal patterns are discernible with regard to social media checkins in Columbus, Ohio?

[6] https://foursquare.com

- In what urban places are social media users actively engaging in the creation of online checkins?
- What light do these social media checkin activities shed on understanding place as a "meaningful location"?

The rest of this chapter is organized as follows: section 3.2 discusses the history of the terms "space" and "place" by exploring theoretical meanings and setting the context for which the terms will be used in this dissertation. In section 3.3 we present a general data collection flow for harvesting geo-social media data. Study area, methods and data of our case study is introduced in section 3.4. We then proceed to demonstrate the results of empirical studies in section 3.5 along with our discussion, followed by summary and discussions on future work in the last section.

3.2 Space and Place Distinction

The concepts of space and place have historically been a popular topic across a vast spectrum of geographic and academic literature. Not surprisingly, the terms do not lend themselves to easy definition. In Berry's (1964) work on the new approaches to the geography of the United States, Berry describes geography studies and the spatial perspective as "the organization of phenomena over space, or the integration of diverse phenomena in place" (page 4). Spaces and places are all around, and we traverse them and experience them everywhere and every day. At the surface, there are some distinctions to be made between the meanings of space and place. Space is often viewed as a more abstract concept than place; spaces evoke the sense of being elusive or distant in nature, taking the form of processes and movements that occur across geographical scales. In contrast, places are often seen as being local and concrete phenomena. Places are often characterized by physical locations on earth's surface that are personal and meaningful. There is a vast literature on space and place, representing great differences in opinion, the most common of which around the argument of which is more important, space or place. While arguments have been taken on both sides, the mainstream view seems to argue that space and place are not separate, but mutually constituted (Wainright & Barnes, 2009). Relph (1976) suggests that space provides the context for places. Building upon this idea, Tuan (1977) characterizes space as movement and place as pauses, for each pause in movement makes it possible for location to be transformed into place. According to Tuan (1977) the concepts of "space" and "place" are dialectical and require each other for definition, for "from the security and stability of place we are aware of the openness, freedom, and threat of space, and vice versa" (page 6). Using this same point of view, Cresswell (2004) further refers to places as spaces that have been made meaningful, arguing: "Space, then, has been seen in distinction to place as a realm without meaning – as a 'fact of life' which, like time produces the basic coordinates of human life. When humans invest meaning in a portion of space and then become attached to it in some way it becomes a place" (page 10).

Agnew (1987) further suggests there are three aspects to place as a "meaningful location": 1. Location, 2. Locale, and 3. Sense of place. First and foremost, in everyday language a place refers to a *location*. Within the physical location of a place there exists the material setting for social relations or the actual shape of the place within which people conduct their lives as individuals, a *locale*. Finally, *sense of place* refers to the subjective and emotional attachment people have to a place. Thus, a place is not just a thing in the world, but a way of seeing, knowing, and understanding the world. One way of understanding modern urban places and the people who interact within them is through the emerging web 2.0 and social media technologies. Goodchild (2007) discusses such online content, usually generated voluntarily by users, has a geographical components as VGI, or volunteered geographic information, which is a highly localized phenomenon characterized by people using and creating their own maps on their own terms, sharing local information and conveying understanding of places to friends and visitors. Since the emergence of VGI, there is an increasing interest and effort in formalizing place in GIScience using data from web 2.0 and social media recently (Goodchild, 2011). In one of these recent efforts, Gao et al. (2014) present a novel approach to harvest crowd-sourced gazetteer entries about places from social media and to conduct high-performance spatial analysis in a cloud-computing environment.

The convergence of social media and GIS (Sui & Goodchild, 2011) provides an opportunity to reconcile the world of space (traditional GIS) and the world of place (social media). Unlike with conventional cities, the trajectories of social media are well documented by the hosting companies; and unlike conventional census data, social media data is defined at individual level, often at very fine spatial and temporal scales. This gives us opportunity to explore human activities in both space and place. For this purpose, we conduct an empirical study in the GIS environment aiming to enrich both the spatial and platial context of social media data, using checkins as an example.

3.3 Geo-Social Media Data Harvest

Social media data, such as Twitter, can be utilized as a new data source in GIScience research since Twitter provides a massive amount of spatiotemporal information about individuals broadcasting their opinions, moods, and activities. Such data usually provide more timely information than that from the traditional sources as social media data could be collected almost immediately after users post their information. Also, the increasing number of social media users equipped with location-based device such as smartphone and tablet produce vast amount of geo-tagged data with spatial footprint. To better understand the space-time dynamics of such social media data, I'd like to build a model to capture the probability of a tweet posted on certain time about certain place. My model includes three modules: data collecting module, data parsing module and data analysis module. Figure 4 below demonstrates the architecture of my model.

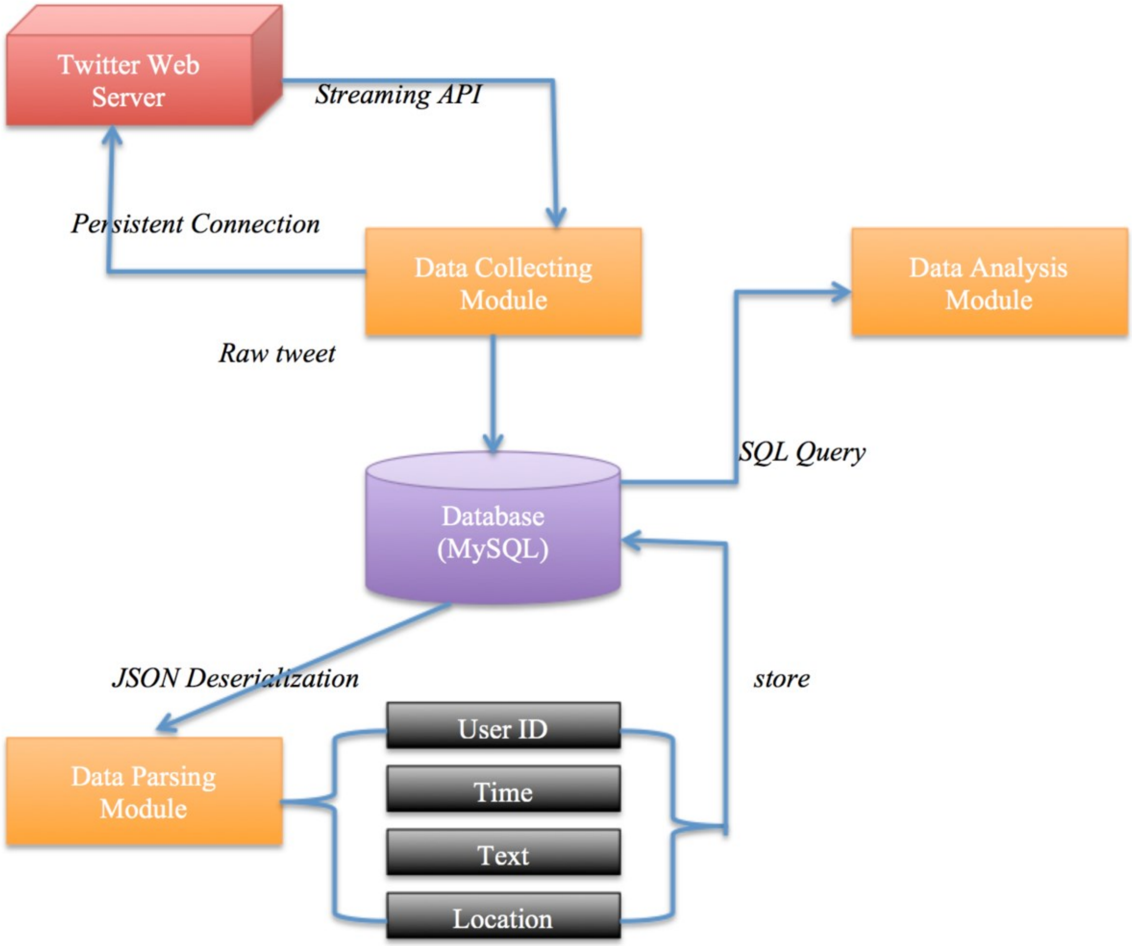


Figure 4 Geo-Social Media Data Collecting Architecture

- Data Collecting Module

To retrieve tweets from Twitter web server, we need to first register a twitter developer account at https://dev.twitter.com. After registration, twitter will generate a group of authentication keys for you, which means you're allowed to obtain their data via a secure connection using these keys. Then we need to establish a connection with the

Twitter web server through its streaming API (Application Programming Interface). Different from the Twitter's search API, which provides a REST (representational state transfer) interface, the method utilized here requires the establishment of a persistent HTTP connection for receiving streaming tweet data. The advantage of using Twitter stream API rather than search API is we're interested in a more general samples of tweets rather than tweets from certain users at certain time. Also, Twitter stream API allows a persistent HTTP connection and once it's connected, we don't need to care about the retrieving module mechanism in detail unless an exception happens such as Internet failures. To account for exception, an error handling mechanism is also adopted to check if the communication channel remains open every time when a new HTTP request is sent and if the number of failed attempts is more than a predefine value, the program will stops and we need to setup a new stream connection.

- Data Parsing Module

We will adopt a database system to store tweets. Once the raw tweets are returned from the server, they are serialized and store in a database (MySQL, for instance). Data parsing module will simultaneously reads the tweets in a timely manner as they are flushed into the database. Since the raw tweets are formatted based on JSON (JavaScript Object Notation) object, we can extract desired information from a text of tweet based on its JSON tag (e.g. user ID, user profile, tweet text, location, time). These new formatted tweets will then be put into another database for further analysis.

Because we are interested in both temporal and spatial information of a tweet and we already have tweet time in our database, the next step in our parsing module is to extract the location information. Unlike the temporal information that can be explicitly extracted, location extraction of tweet is often difficult and our proposed location parsing module are threefold. First of all, among all posted tweets, about 2.5% of tweets are available with location information either in the “geo” tag or the “place” tag from the raw tweet. So we’ll firstly filter those 2.5% tweets with geo-tag and then extract the location directly from itself. If the location information is expressed as a pair of [latitude, longitude], we need to also conduct a reverse geocoding in order to retrieve the place information from a point location. A second step, we can create a filter for those 97.5% non-georeferencing tweets. The key words will be the name of the location. For example: Let X as the set of our n desired locations and

X_1=Columbus,

X_2=Cleveland,

X_3=Cincinnati,

…

X_n=New York.

Then we will write a SQL command to query the tweets content and find out those tweets contains a particular name of the location. Then we will assign the name of the location to this tweet as its location information for further analysis. Last but not the least, there’re recent efforts from GIScience and computer science to infer the location of a non-georeferencing tweet. For example, Cheng et al. (2010) propose a probabilistic framework to infer the location of tweet based on its twitter user and its content. If necessary, we could adopt Cheng’s method to retrieve more tweets to enrich our location information.

- Data Analysis Module

Once we have our tweets formalized in a database, we can start our analysis module. Our ultimate purpose is to calculate the probability of a tweet at certain time and certain place. This could be expressed using the formula:

$P(X(T)=X_i)=$

$$\sum_{t0}^{tn}\frac{(N(X=\mathrm{Xi}))}{N(X)}\ where\ N(X) means\ the\ number\ of\ tweets\ and\ U\{t0, t1, t2, \ldots tn\} = T$$

Here is the interpretation of this formula. If we want to get the probability of the occurrence of tweets about a place X_i within a period of time T, we need to first count all the tweets with location information equals X_i from time t0 till time tn and we take it as $N(X = \mathrm{Xi})$. We use t0 to represent the starting time of T and tn as the ending time of T then all the tweets in our database will be count as N(X). Then we use N(X=Xi)/N(X) and we will get the probability.

3.4 Study Area, Data and Methodology

Our case study is conducted at a regional scale in order to capture the space and place information at a greater detail. Columbus, Ohio, a capital city in the Midwestern U.S., was chosen as our study region. This city is home to many business headquarters and a variety of industries. For this reason, it is frequently referred to as a “test city” (Hunker, 2002) for new consumer goods and thus makes it a good site for our study.

Social media website such as Foursquare and Gowalla let users “check in” to a place when they’re there, tell friends where they are and track the history of where they’ve been and who they’ve been there with. To begin our study, we first acquire a collection of “checkins”. Since personal checkin information about location sharing services on Foursquare and Gowalla is typically restricted to share in the user’s private social network, it’s also not publicly available via location sharing service websites such as Foursquare and Gowalla. We then take an approach in which we sample location sharing status updates from the public available Twitter feed. Twitter status messages support the inclusion of geo-tags with latitude and longitude information as well as support third- party location sharing services such as Foursquare and Gowalla if a user agrees to opt in to share their checkin status on twitter account. We thus monitor Twitter’s streaming API[7] (Application Programming Interface) which represent around 10% of the whole twitter database and retrieve users who post geo-tagged status updates.

The checkin collecting program ran from late February 2010 to late January 2011, collecting a total of 1850 users and 50,662 unique checkins with geographic location situated in Columbus, Ohio. Each checkin is stored in a local database with the format: checkin (userID, tweetID) = {userID, tweetID, text, location, time}. An example checkin in our study area is presented in Table 2.

Table 2 Example checkin record in database

userID	**tweetID**	**text**	**latitude**	**longitude**	**time**
154463 86	1233595 3358	Heat sheet in hand... (@ Hoover Invitational) http://4sq.com/d1P63w	40.118	-82.885	2010-04-17 10:58:54

To explore the checkin information in space we first plotted the locations of checkins on a map using ESRI ArcGIS 10.1, a leading GIS platform and were overlaid

[7] https://dev.twitter.com/docs/api/streaming

within the boundary of Franklin County in Central Columbus, Ohio. To further explore the spatial distribution of checkins, we apply methods from Exploratory Spatial Data Analysis (ESDA), which is a set of techniques aimed at describing and visualizing spatial distributions and identifying hot spots and clusters in space (Bailey and Gatrell 1995). More specifically, to estimate the number of checkins occurrences per unit area, we performed a kernel density analysis of Columbus area using checkin data. Kernel density is a popular method of estimating the intensity of points by creating a smooth surface using a bivariate probability density function (Bailey and Gatrell 1995). The kernel estimator is defined as:

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-xi}{h}\right) \qquad [1]$$

where n is the total number of points, h is the bandwidth that determines the amount of smoothing, K is the kernel function, x is the location of estimation, and x_i is a known point location. To further test whether checkins at different census tracts within Columbus, Ohio tend to be clustered together in space or dispersed, we apply Global Moran's I and Local Moran's I which are the most commonly used test statistics for spatial autocorrelation (Anselin, 1995).

$$I=\frac{N}{\sum i \sum j \omega_{ij}} \frac{\sum i \sum j \omega_{ij}(X_i-\bar{x})(X_j-\bar{x})}{\sum i (X_i-\bar{x})^2} \qquad [2]$$

The information about place venues that we use to relate to different checkins came from InfoUSA, a company that provides business data and marketing services. Specifically, the raw data in the Columbus area has a total of 44,097 records, containing all types of businesses venues. These were further grouped into 8 categories by employing the NAICS (North American Industry Classification System) codes and those venues receiving no checkins were eliminated in our studies. We choose to enrich the platial information of a checkin in terms of its nearest neighbour business venue. In order to find out the nearest business venue, point distance from ArcGIS 10.1 is applied to calculate the distance from each checkin point in one feature class to all the business venues points within a given search radius 25 meters. The business venue in the nearest distance from the output table will be matched to the checkin point. An illustration of point distance function is presented in Figure 5.

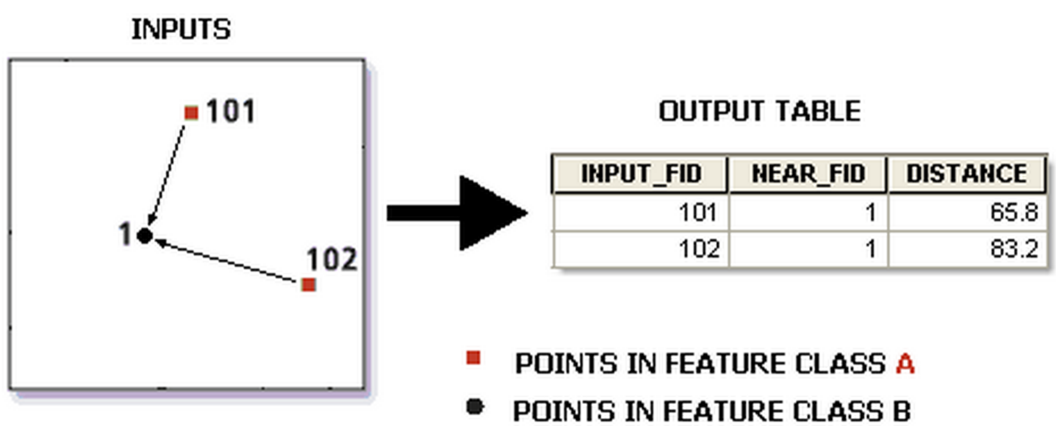


Figure 5 Illustration of Point Distance Function[8]

[8] Figure from ESRI

3.5 Findings

- Checkins in Space

In this section, we begin our study of social media location sharing services with an investigation of the spatial and temporal distribution of the checkins. First, we plot the locations of our over fifty thousand checkins shown in Figure 6. Checkin locations roughly delineate the major street network and the urban area within the Columbus area at a very good resolution.

To explore the spatial density of checkins, we create a surface map using kernel density estimation (Figure 7). There are two parameters in kernel density estimation: kernel bandwidth and cell size. The kernel was set as 10 km and the cell size was set as 250 meter by 250 meter given the size of our study area. The kernel bandwidth of 10 km is a compromise between a map that is too smooth to interpret and one that is too noisy to interpret. The cell size of 250 meter was used to show fine detail of an urban area. Downtown Columbus and The Ohio State University main campus area have the highest checkin density and it gradually decreases in the surrounding areas. Port Columbus international airport to the northeast of downtown is also well covered with checkin footprints.

To further investigate the spatial autocorrelation of the checkins pattern across Columbus, global and local Moran's I is used here. Global Moran's I is 0.22 at a statistically significant level, which indicates a positive autocorrelation of the checkin activity across the city. In other words, similar density of checkin activities in surrounding census tracts is identified for a focal spatial unit. The neighbourhood effect is generally held across the study region. As we can see in Figure 8, the cluster with a high-high relationship is the area where downtown and Ohio State University main campus school district located. The industrial belt south of the downtown area is where the low-low cluster presents, which partially reveals the industrial decay. An interesting finding is to take a look at the outlier. For example, a high-low outlier is the census tract where a popular shopping centre (Easton Town Shopping Centre) is located. In other words, this shopping centre overwhelms its vicinity on checkin activities reported through social media.

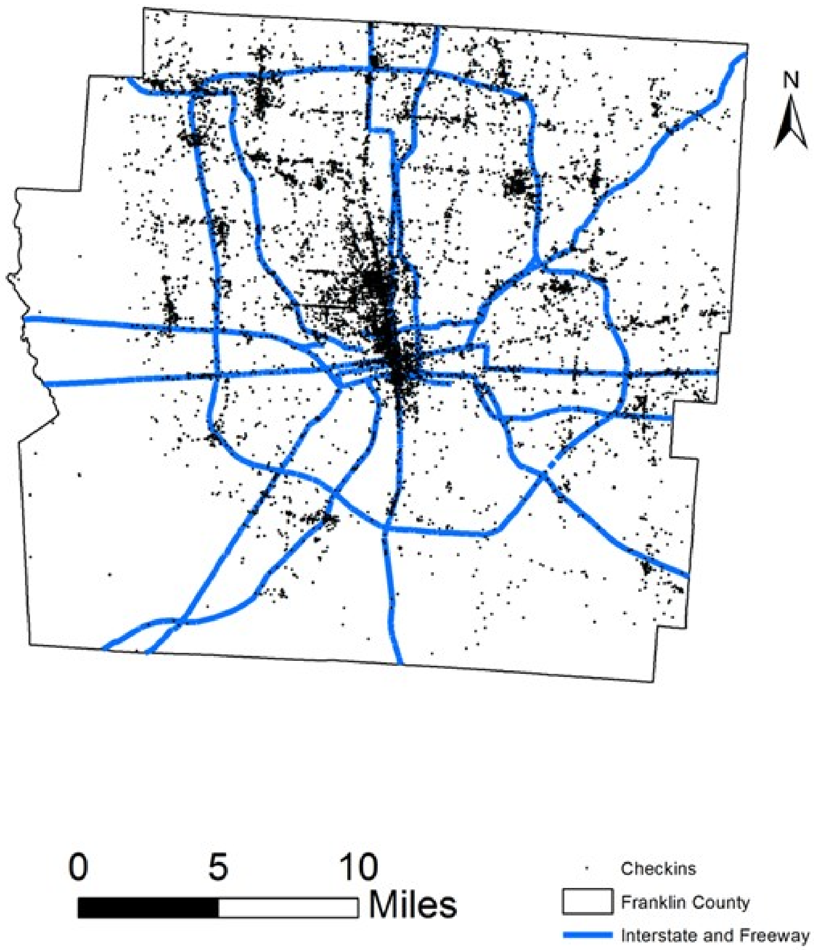


Figure 6. Overview of checkins in Franklin County, Ohio

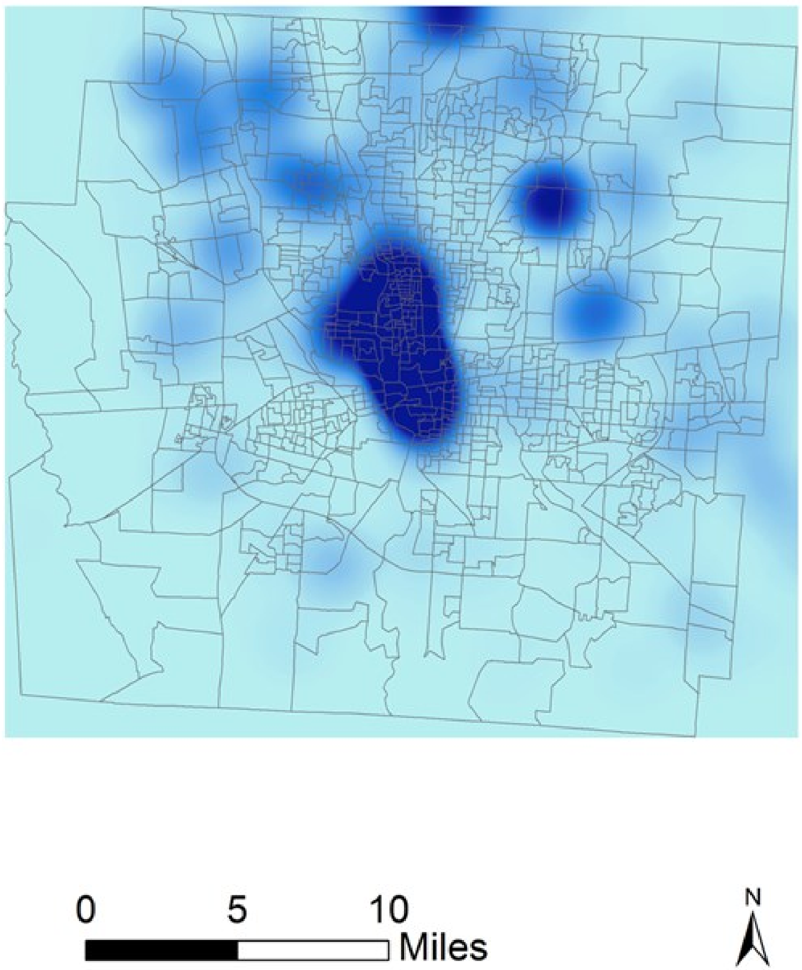


Figure 7 Kernel density of checkins in Franklin County, Ohio

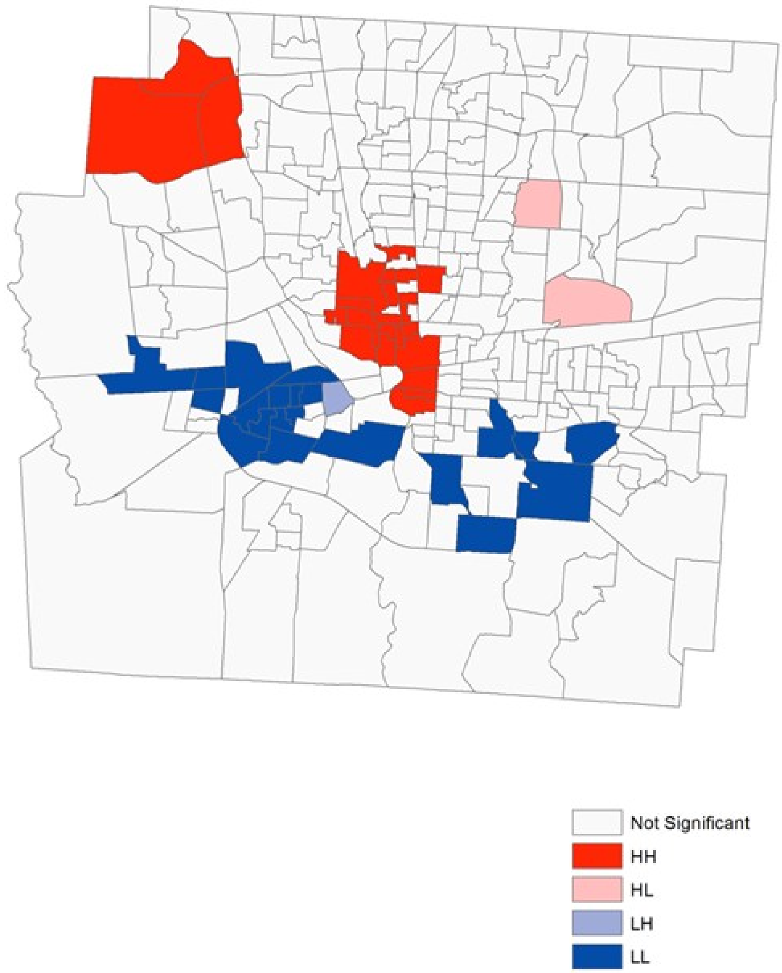


Figure 8 Local Moran’s I cluster map of checkins

- Checkins in Place

Over 44,000 place venues from InfoUSA have been geocoded in ArcGIS environment. In Figure 9 we present the spatial distribution of place venues in the Columbus area categorized using eight different colors. Our point of interesting matching of checkin and venue shows that among over 44,000 venues, 10,847 venues are assigned as food venues and there are a total of 13,617 checkins belonging to this category. Services venues receive the second highest amount of checkins (i.e., 10,542) and retail shops attract over 5,000 checkins.

Table 3 Place venues in Columbus from infoUSA

Category	Number of Venues	Number of Checkins
Food	10,847	13,617
Retail Shop	5,590	5,334
Services	18,767	10,542
Public Administration	2,442	1,969
Transportation	1,611	1,118
Industry	1,426	865
Outdoor Activity	694	299
Others	2,720	16,918

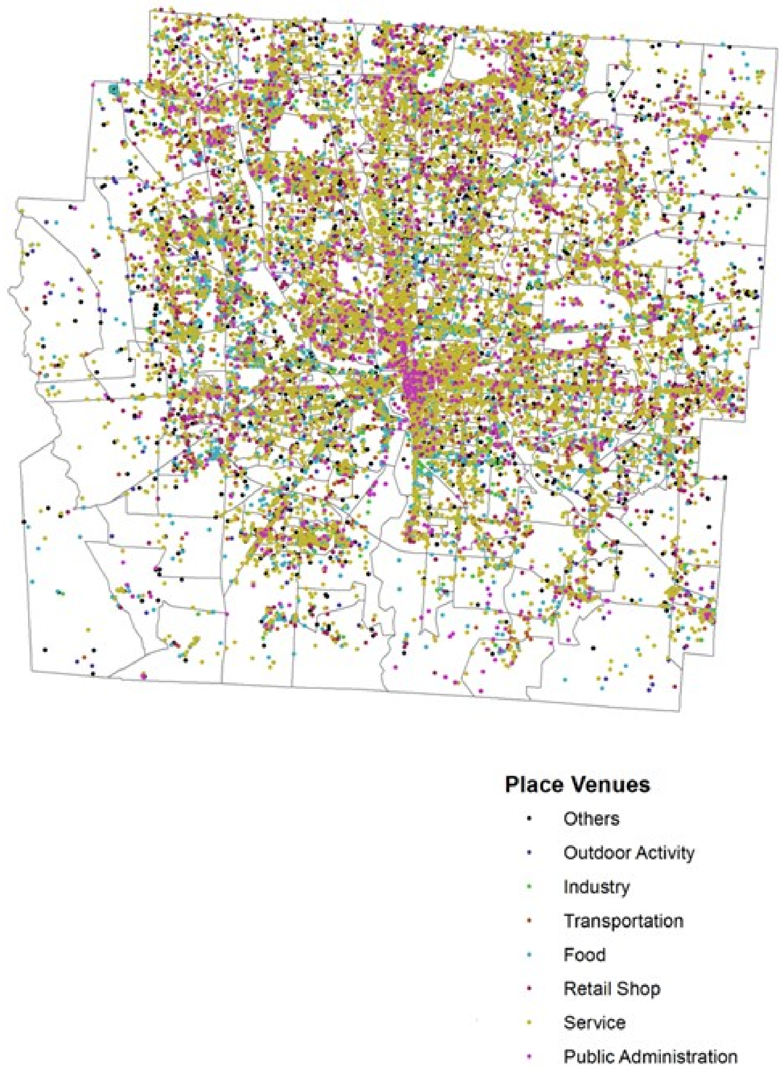


Figure 9 Place venues in Columbus in different categories

To examine the representation of the places on social media in Columbus, we then conduct an analysis on Foursquare website to find out business venues that receive most checkins by August 2014. In Foursquare, both single users and business organizations can create a venue that is tied to any specific group, business, company, or location. With few exceptions, every business venue is tied to a specific geographic location. For instance, a page can be created for a single restaurant, a university, or an entire neighborhood. Also, each business venue has information about how many users have visited and how many checkins received. These business venues pages on Foursquare also contain information about to what degree people like a place using a rating system, in which 10 are the highest score and 1 to be lowest one. We believe such information is a good implication of people's sense of a place. In order to evaluate each place's location, locale and people's sense of place, a table was compiled by searching Foursquare by setting the geographic filter as "Columbus, Ohio". Based on this search, we presented top 25 venues with most checkins number as a subset of representative places from social media.

For each place, an address is recorded as well as the number of population who has visited, the type of the venue, number of total checkins, and the score of the "like" as sense of place. This database is documented in Table 4 ranked by the number of checkins. As of August 2014, 25 top business venues represent 216,237 foursquare users and 774,886 checkins. From the locale perspective, this list includes Port Columbus International Airport at Number 1 (198,074 checkins) as well as Easton Town Center (68,086), Nationwide Arena (39,083) and Polaris Fashion Place (37,093). In terms of geographic location, as hypothesized most of these business venues were attached to places in downtown areas (Figure 10). Most venues are concentrated on the North-South

Axis across the downtown core and this pattern matches up with the overall checkins distribution presented in the previous section.

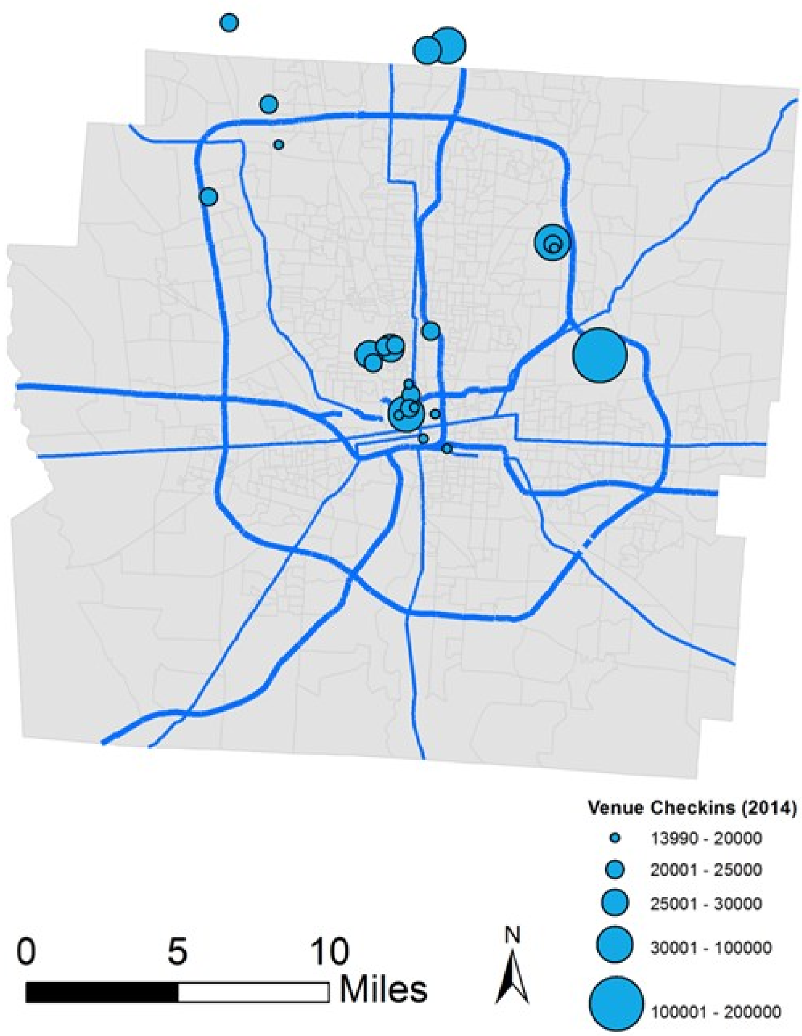


Figure 10 Top 25 venues in Columbus from Foursquare

Table 4 Top 25 place venues in Columbus from Foursquare

Venue	Visitor	Checkins	Type	Like (1 –10)
Port Columbus International Airport	50,720	198,074	Airport	7.5
Easton Town Center	16,893	68,086	Mall	9.0
Nationwide Arena	12,246	39,083	Stadium	9.2
Polaris Fashion Place	11,827	37,093	Mall	7.8
JPMorgan Chase	1,264	28,524	Office	N/A
Recreation & Physical Activity Center	3,075	26,710	Gym	8.8
AMC Lennox Town Center 24	7,867	25,382	Theater	7.7
The Mall at Tuttle Crossing	7,276	24,856	Mall	7.2
LA Fitness	1,221	23,766	Gym	6.2
Ohio Stadium	10,903	23,602	Stadium	9.4
Union Café	5,072	23,262	Gay Bar	8.3
Columbus Zoo & Aquarium	12,297	23,008	Zoo	9.5
Target	5,739	22,329	Store	6.5
North Market	9,770	21,749	Market	9.4
Columbus Crew Stadium	7,809	20,954	Stadium	9.3
The Ohio State University	6,110	20,590	Education	N/A
AMC Easton Town Center 30	7,927	20,238	Theater	8.4
Columbus State Community College	2,172	19,385	Education	N/A
Huntington Park	8,209	18,725	Stadium	9.4
Nationwide Children’s Hospital	2,299	15,887	Hospital	N/A
Fado Irish Pub & Restaurant	4,026	15,753	Food	8.0
Bodega	4,320	14,989	Bar	8.5
Dirty Frank’s Hot Dog Palace	5,757	14,763	Food	9.3
Greater Columbus Convention Center	7,185	14,088	Convention Center	8.1
Whole Foods Market	4,253	13,990	Grocery	8.9

3.6 Discussions and Conclusion

- Bring Time in Consideration

Besides exploring the spatial distribution of checkins and the platial context of each checkin, there is further potential to understanding the checkins. For example, who are the checkin users and how often do they checkin.

In Figure 11, we plot the number of checkins generated by each user versus the number of these users. As the figure shows, most users contribute less than 10 checkins in Franklin County, Ohio during our 11-month study period, though there are nearly one-tenth the users (143) contributing more than 100 checkins, with 911 checkins being the most active user in our study. The real time data collection made possible by social media should be accessed at its full potential. Figure 12 provide a general overview of how Foursquare users interact with places in our study area. As depicted, people check in frequently at around noon and in the evening. Most activities occur between 6am and 23pm, with two peaks at around 12pm and 6pm. This is probably due to a fact that these are periods when people often go for lunch and dinner, so most checkins at these two peaks might be related to restaurants and food. Further study can be conducted to study traffic patterns over 24 hours in a day and over 7 days in a week of certain place venues, such as restaurants or food stores. This investigation can illustrate more dynamic details of people's activity in space and place. Such study could be applied for retail analysis such as the creation of more convenient store hours as business retailers' hours are limited. How to make complete use of social media for balancing the tradeoff between the needs of customers and the cost of store operation would be a potential trend for future studies.

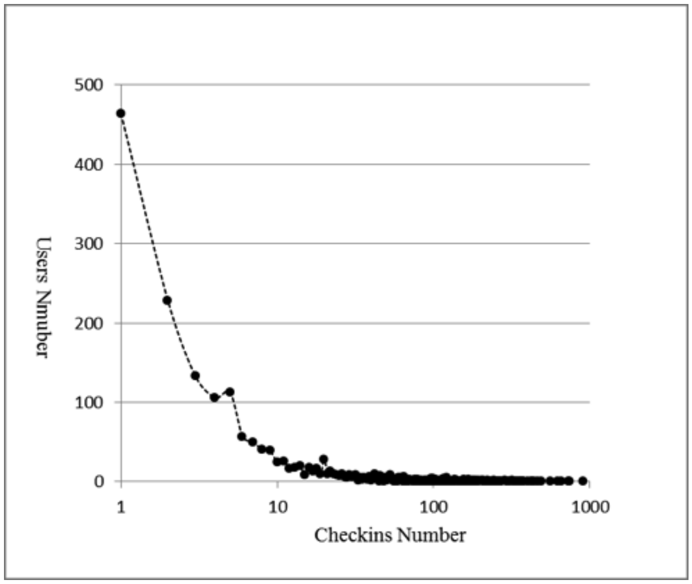


Figure 11 User checkin statistics

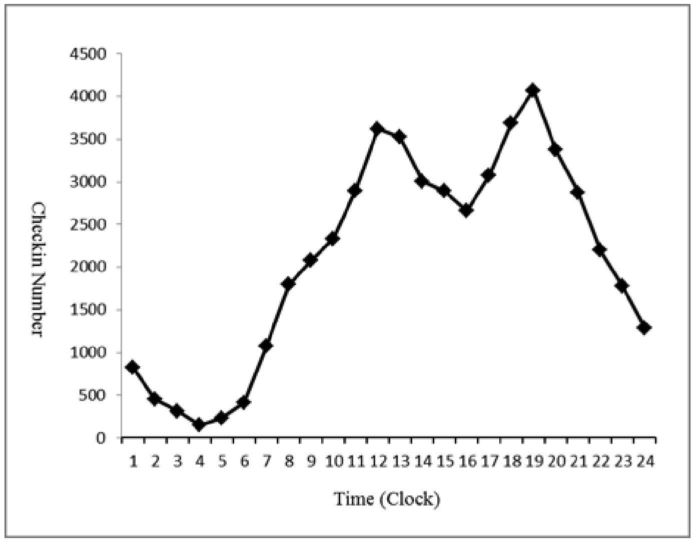
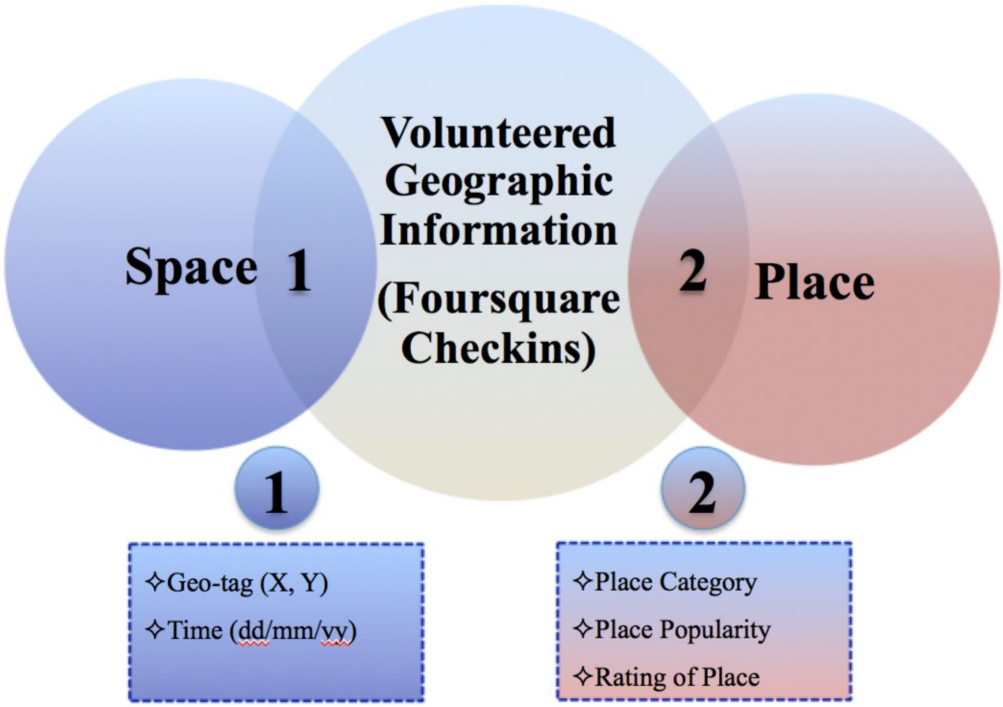


Figure 12 Daily temporal distribution of checkins

- More on Place

By examining the popular Foursquare venues, we find that more than 50 percent of popular venues in our database are located in downtown area. The abundance of Foursquare venues and checkins around the core areas of downtown Columbus provide some important insights. First of all, areas in Columbus downtown are newly revitalized thus have become unique and sticky places (Knox, 2011). They are not only facilitating interaction among people in person, but also online in social media as well. Also, such large amount of visits on Foursquare in comparison to other areas in the city indicate that these areas have created a *"Sense of place"* (Agnew, 1987), with people feeling compelled to express their identification with the area among friends online using Foursquare's rating system as we present in Table 3. In addition, the top-25 popular venues reported in Table 3 are mainly hangouts spots, such as bars, restaurants, stadium, etc. From the "*locale*" perspective, these could be interpreted as popular physical location existing as the material setting for people's social interactions and the place within which people conduct their lives (Agnew, 1987). Last but not least, it is worth noting that, in reality, not everyone uses the location based social media service, such as Foursquare, or has it available to them due to the intra-urban digital divide (Baginski et al., 2014). Such finding indicates that places of downtown are the most popular among young professionals and college students, a demographic assumed to be the primary users of the most up-to date communication technology. Other places in downtown, such as German Village, may be just as popular but may include an older demographic less likely to express themselves online. We need to consider such biased issue of social media data in further research.

In retrospect, this study explores the urban dynamic in Columbus, Ohio using data harvest from social media. Such checkin data contains geo-tag, which could be used to pinpoint each checkin within the geographic region of Columbus. By aggregating these in space, we are able to present the spatial distribution of these checkins as well as to identify areas receiving high volume of checkins than others. The temporal information of each checkin further tells us when the checkin is recorded thus we are able to have a general sense of the temporal distribution of urban activity in 24 hours. This example shows that geo-tagged social media data such as checkins normally contain spatial and temporal signal that we could use to explore in space, as denoted the intersection "1" in figure 13. Meanwhile, our study also shows that checkins contains elements (intersection



"2" in figure 13) that could be used to characterize place such as place name, place category, people's checkin numbers on certain places, and people's rating of place. By synthesizing such elements we have a better understanding of the urban dynamic about different place in this city.

Figure 13 Framework to Explore Urban Dynamic in Space and Place

Chapter 4: Use VGI to Study the Relationship between Individual Food Environment and Food Choice[9]

Chapter 3 demonstrates how geo-social media could be used to explore the pulse of a city in space and place. In this chapter, we tackle the challenge of using VGI to study people's activity in related to their surrounding places as a place-based measure would generalize attributes that vary among people. Specifically, we investigate people's food related activity harvested from twitter and report a quantitative assessment of their food choice relationship between their exposed food environments to food vendors.

4.1 Introduction

Access to nutritious food is imperative to physical well-being and quality of life. Failing to consume healthful food on a regular basis can lead to a series of adverse health outcomes, including obesity, diabetes and cardiovascular diseases (Shaw, 2006). Researchers have noted that access to nutritious food is influenced by both spatial and non-spatial mediators that result in inequality of access across communities. These barriers to healthful food choices include geographic, economic, informational, and cultural aspects (McEntee & Agyeman, 2010). To date, the majority of the literature simply focuses on geographic access to nutritious food in relation to socio-economic status (SES). These studies have taken a predominantly statistical approach to examining

[9] Portions of this chapter have been previously published in Chen and Yang (2014).

if correlations exist between SES variables and food access though regression models. Not surprisingly, many studies found a positive correlation between low SES and limited access to quality food in selected local regions (Glanz et al., 2007; Moore & Diez Roux, 2006) as well as in U.S. nationwide studies (Powell et al., 2007), and some other studies failed to identify this correlation (Guy & David, 2004).

An overlooked facet in this spatial disparity of food access is the causality linked to food choices at the individual level. Individuals tend to make diverse food purchasing and dining choices, including where, when, how, and which types of food to acquire. The role of food access in shaping food choices cannot be completely understood from generalized regional studies. What must be assessed is the impact of the quality of an individual's food environment on his or her food purchasing choices. Although studies have previously identified among other factors the availability of nearby grocery stores have played a significant role in influencing food buying practices (Walker, et al., 2011; Walker, et al., 2012), these studies suffer from the following limitations: (1) the sampling size is very limited due to the time-consuming process of collecting individual samples and (2) studies of individuals' dietary choices are based on the home locations while overlooking the effects of mobility on procuring food (Kestens, et al., 2010; McKinnon, et al., 2009). At the individual level, the deficiency in food procurement may be due to a scarcity of healthful food retailers in the individuals' vicinity.

An unexplored direction is to measure if people's preference for food is elicited by exposure to a particular type of food environment in real time, such as a lack of quality grocery stores nearby that would prevent them from making nutritious food choices. This finding would also offer compelling evidence for shaping strategies to improve the health of communities.

The traditional measure of food access based on location of food outlets has considerable limitations for understanding the causality of food choices (Wrigley, et al., 2002); on the other hand, individual-level analysis has always posed a daunting challenge for social scientists because of the tedious nature of collecting a considerable number of representative samples. To fill this gap, this study takes an innovative approach to collecting individual data about food-related activities from Volunteered Geographic Information (VGI) via social media, or specifically, “tweets” (messages sent on Twitter). Tweets not only contain self-reported information about life experiences and activities but are also tagged with accurate location and time information. They provide an ideal solution for measuring exposure to the food environment at the exact time when the user tweets. This measure, as a representative of individual food access, is associated with users’ diet choices conveyed in their tweets. We then analyzed them to determine if there was a significant association between their surrounding food environment and the quality of their particular food choices. When groups of Twitter users who shop in grocery stores were compared to those who dine at fast food restaurants, it was revealed that the prevalence of nearby fresh produce grocery stores may significantly influence individuals to make healthful food choices. Conversely, density of fast food restaurants in an individual’s immediate area may not discourage healthful food choices.

To the authors’ knowledge, this is the first time social media has been adopted for the evaluation of individuals’ food-related activities that are influenced by one’s local food environment. This study has great potential for informing health scientists about the importance of using social media to measure individual behavior and for informing community advocates and stakeholders of the significant role quality food retailers can play in changing the dietary practices of area residents. Our study also shows the capability of using VGI to study food accessibility from a holistic view by linking people-based and place based together.

4.2 Related Work

The problem of food access can be interpreted in many different ways. It not only involves geographic access to food but also pertains to affordability, availability of culturally appropriate food, and knowledge about nutrition (McEntee & Agyeman, 2010; Shaw, 2006). Beyond these descriptions, the majority of studies approach the problem from a rather geographic perspective, where the metaphor "food desert" is employed to illustrate areas devoid of safe, affordable grocery stores with extensive arrays of quality food items. Such grocery stores which offer fresh fruits and vegetables are termed "green retailers" (Wrigley, et al., 2002). A consequence of a lack of such retailers in urban food deserts is an infiltration of poor-quality food outlets, such as convenience stores and fast food restaurants, where only packaged, processed, and energy-dense food is provided (Drewnowski & Specter, 2004).

Reducing and eliminating food deserts has been the goal of stakeholders and government agencies around the world ever since the term was first coined by the U.K.'s Nutrition Task Force's Low Income Project Team in 1996 (Reisig & Hobbiss, 2000). Since then, it has become a topic of interest in a Congressional report compiled by the

U.S. Department of Agriculture (USDA) in 2009 (USDA, 2009). A method for evaluating the formation of food deserts is needed before a solution can be found. To date, there is a lack of consensus on methods used for differentiating high/low access areas and demarcating food deserts. With respect to measuring food access, the basic rationale differs from either a place-based perspective or an individual-based perspective.

A place-based perspective examines the built environment in which people acquire food in multiple contexts, such as food stores, restaurants, worksites, and schools (McKinnon, et al., 2009). A typical place-based approach measures the level of geographic access in terms of the diversity, proximity and variety of food outlets in a predefined geographic unit, such as a neighborhood, census tract, census block, zip code zone, county, or state (Apparicio, et al., 2007). The geographic unit serves as a container for summarizing the spatial attributes of food outlets in terms of total number (Berg & Murdoch, 2008; Powell, et al., 2007), number relative to population (Ball, et al., 2009), average number within a distance of the centroid of the unit (Larsen & Gilliland, 2008; Raja, et al., 2008), and distance to the nearest store from the centroid (Larsen & Gilliland, 2008; Pearce, et al., 2006). Disparity of access has been identified across geographic units, compounded with regression models for determining if isolation from a quality food environment is correlated with racial composition or with socio-economical deprivation of districts. Another place-based measure, which relies on the emergence of Geographic Information Systems (GIS), takes into account the spatial distribution of food resources by creating a circularly buffered zone or a network-shaped zone around the focal location of food outlets (Algert, et al., 2006; Chen & Clark, 2013; Clarke, et al., 2002). The creation of the buffer is grounded in the perception that sustainable neighborhoods should have sufficient access to nutritious food within a walkable distance. Areas outside the food access zones are considered underserved and are in need of policy intervention. In spite of the container measure and buffer measure, other place-based measures explore emerging geographic analysis methods by examining the spatial relationship between food supply and demand. Examples of these methods include various accessibility models (Helling & Sawicki, 2003; Paez, et al., 2010) and spatial interaction models (Clarke, et al., 2002).

The place-based approach is limited by the fact that it generalizes individual attributes, such as gender, age, ethnicity, and other demographic aspects that vary widely among people (Miller, 2007). Because people's life experiences vary, their preferences for food and knowledge of nutrition are difficult to infer from aggregated census data. In order to capture these distinctions, the individual- based approach can be adopted by focusing directly on the food choices of people grouped by neighborhood, age, gender, and social roles. Importantly the foci of these individual level studies include general households (Wrigley, et al., 2002) or selected groups such as low-income rural residents (Smith & Morton, 2009), women (Ball, et al., 2006; Laraia, et al., 2004), and male adolescents (Jago, et al., 2006). Once the population of interest is defined, standardized assessment tools in the form of videotaped interviews or paper-based questionnaires are customized and distributed to the specific population. The proposed survey questions are not only limited to accessibility to food but are more focused on the respondents' financial status, nutritional knowledge, health concerns, food purchasing and consumption behaviors, and the neighborhood food environment that shapes their daily food choices. Geographic access is also explored through various methods for measuring foodscapes at the individual level by referring to the home address of respondents (Ball, et al., 2009; Jago, et al., 2006; Jeffery, et al., 2006; Laraia, et al., 2004; Wrigley, et al., 2002).

The majority of previous individual-based studies are flawed in two ways. First, home addresses are fixed locations that only represent a limited range of activity. People's daily movements can extend beyond their households and expose them to other food environments, such as cafeterias at employment sites or schools (Kestens, et al., 2010). Their mobility not only changes the physical environment that limits the choices of food resources but also introduces food cues, such as advertisements and interpersonal communications that may elicit an increased craving for certain foods (Ferriday & Brunstrom, 2008). In addition, individuals' food procurement is not only spatially but also temporally constrained. As the operating hours of food stores are invariably limited, those whose discretionary time conflicts with store hours will find it difficult to access food, eventually severely limiting the quality of their food choices (Chen & Clark, 2013). Previous food studies, however, have overlooked this temporal inaccessibility to food stores and the contraction of activity space impinged by time.

The recent trend centering on individuals examines exposure to one's local food environment in terms of availability or density of food resources attainable in one's daily travels. The goal of these studies is to identify the mechanism by which environmental influences translate into individual food-related choices and to highlight the SES indicators that contribute to these choices. Previous studies have relied exclusively on techniques that trace individual movement: travel diary surveys (Kestens, et al., 2010) and GPS tracking (Chaix, et al., 2013). Travel diaries are collected in the form of paper-based or online surveys from individual participants that document their activities over a specified span of time. They include questions to determine an individual's attitudes towards their search for food sources, which aid in behavioral and psychological analysis. However, the design, collection, processing, and analysis of individual travel diaries constitute a tedious and labor intensive process. The level of accuracy and precision of responses regarding the location and time of food-related activities also must be verified. GPS tracking allows for a more automated and accurate data collection process; however, the lack of contextual information poses challenges to categorizing GPS tracks into separate activities and the segmental trips between them (Chaix, et al., 2013). The geographically referenced individual data with excessive spatiotemporal details has been criticized for intruding on individual privacy (Kwan, et al., 2004). An improved approach is to combine GPS tracking with a follow-up activity survey. Unfortunately, previous studies using this approach invariably limited the time span of the survey to only one day (Chaix, et al., 2013).

Recent studies in applied geography have made use of Web 2.0, a concept coined to describe user-generated web contents in a bottom-up approach via social media, such as Twitter, Facebook, Flickr, and Googleþ (Arribas-Bel, 2014). These self-reported contents, known as VGI, offer an alternative approach to keeping track of food-related activities from individual users' dietary preferences and shared personal experiences.

Social media is becoming a new instrument for analyzing individual awareness of health information and prevalence of diseases (Robillard, et al., 2013). Emerging as the leading platform of social media, Twitter has established an active user base of 5.5 millions (Twitter, 2013), and this number has grown at an unprecedented rate in recent years with the increasing popularity of web-enabled mobile devices. In the spectrum of health studies, Twitter has been used to explore the spatial distribution and temporal variation of health-related subjects, including flu epidemics (Chew & Eysenbach, 2010), cholera outbreaks (Ehrenberg, 2012), and causes of dementia (Robillard, et al.,2013). To the authors' knowledge, this study is the first to engage Twitter for measuring the impact of the food environment on individual eating behavior.

Twitter messages, or tweets, serve as an ideal instrument for examining individual food access for several reasons: (1) tweets are geo-tagged in space and positioned in time, and with this precisely defined spatiotemporal information, the dynamic food environment to which the user is exposed can be retrieved with a high degree of accuracy (Arribas-Bel, 2014); (2) tweets entail user- generated individual activities, and individuals' particular food choices can be reflected by their content, allowing for food themed analyses; (3) tweets are voluntarily disseminated by individual users and are open to the public. This self-reported mechanism overcomes the privacy concerns of acquiring geocoded individual information.

4.3 Methodology

Unlike previous social media studies that encompass a relatively large geographic area, our case study was demarcated at a regional scale in order to evaluate the food environment in greater detail. Columbus, Ohio's capital city in the Midwest U.S. was adopted as the study region. According to the U.S. 2010 census, Columbus has an estimated population of 787,033, making it the largest city in Ohio. For this reason, it frequently functions as a "test city" for new consumer goods and attracts diverse industries (Hunker, 2000). The city is home to many business headquarters and fast food restaurants, including Wendy's, White Castle, and Bob Evans restaurants. The diversity of the food industry and a burgeoning population in Columbus make it an accurate representation of U.S. cities. We collected two sets of data within Franklin County, the largest county in Columbus: Tweets with verified content that include information about a current or upcoming food choice and food outlets, including quality grocery stores and fast food restaurants in the region. These two datasets were correlated during analysis to examine the relationship between Twitter users' food choices and their exposure to the local food environment.

Raw tweets that were (1) disclosed as publicly available and (2) tagged with the sender's location over the span of five weekdays (from Dec.2 2013 to Dec.6 2013) were streamed directly from Twitter using an open source Python library. These tweets were geocoded using ESRI ArcMap 10.2, a leading GIS platform, and were overlaid within the boundary of Franklin County in Central Columbus, Ohio. Eventually, a total of 81,543 tweets within the boundary were collected for further verification. In order to retrieve tweets that contained only food-related activities, they were narrowed down by using key word searches. Key words were meant to capture two very different sets of food sources available in this region: names of grocery store chains and individually owned stores selling fresh fruits and vegetables, known as “green retailers” (Wrigley, et al., 2002), such as Kroger’s, Giant Eagle, Walmart, Meijer, Trader Joe’s, Save-A-Lot, etc., and names of fast food restaurants listed in the local Yellow Pages, such as McDonald’s, KFC, Subway, White Castle, Burger King, Taco Bell, Chipotle, etc. Of these selected tweets those with erroneous information or themes unrelated to the following four categories were excluded: (1) words that have double meanings, e.g., “save a lot” referring to a promotion instead of a store brand; (2) statements and opinions irrelevant to food choices (while tweets expressing desires to make a trip to a grocery or fast food restaurant were kept), e.g., “Meijer Christmas lights commercial this year seriously brings me to tears; ” (3) activities that occurred in the past or were scheduled far into the future (while near future activities were kept), e.g., “Just came from Walmart shopping for the kids; ” (4) activities occurring in a place other than the food venue or activities not conducted by the Twitter user, e.g., “Someone brought me Subway and Pepsi!” To ensure the validity of the data, tweets with a vague context were judged at the researchers’ discretion. After irrelevant tweets were excluded, the selected grocery shopping- themed tweets (termed “healthful tweets”, n = 61) and fast food dining themed tweets (termed “unhealthful tweets”, n = 296) were geocoded in the study region using their disclosed latitude and longitude coordinates. Table 5 shows examples of these tweets including the sending time and concurrent coordinates. Figure 14 (a) shows their geographic distribution within the study region.

Table 5 Sample tweets indicating food-related activities in the study region

Date	Time	Content	Latitude	Longitude
12/2/2013	10:43	I'm at @Walmart Supercenter (Whitehall, OH)	39.9543	-82.9004
12/2/2013	11:57	I can't beleive Giant Eagle has such poor customer service. Only a few registers open around lunch time and the lines are long!	40.0322	-82.9083
12/2/2013	12:50	Someone come(s) with me to Chipotle.	39.8742	-83.0486
12/2/2013	15:42	Sitting here at five guys alone.	39.9886	-83.0023
12/2/2013	18:25	Only thing getting me through studying is the food I'm going to get at Kroger.	39.9974	-83.0091
12/4/2013	18:51	Pulled my hair straightener, toothpaste, Tupperware out of my purse looking for my wallet in the checkout line at Kroger.	40.089	-82.8246
12/5/2013	12:14	Taco Bell for the third day in a row.	40.0318	-83.1417

The information about green retailers and fast food restaurants came from InfoUSA, a company that provides business data and marketing services. Specifically, the raw data containing all types of businesses in the study region were narrowed down by employing the NAICS (North American Industry Classification System) codes of supercenters, supermarkets/convenience stores, and fast food restaurants. Grocery retailers with fewer than five employees were excluded to ensure that all the stores mentioned in the study carried a wide range of fresh produce. Fast food restaurants were further distinguished from full-service restaurants and untraditional food retailers, such as snack shops, bakeries, bars, etc. that may also carry unhealthful food items. These selected green retailers (n = 118) and fast food restaurants (n = 2106) were geocoded in ArcMap, as shown in Figure 14 (b).

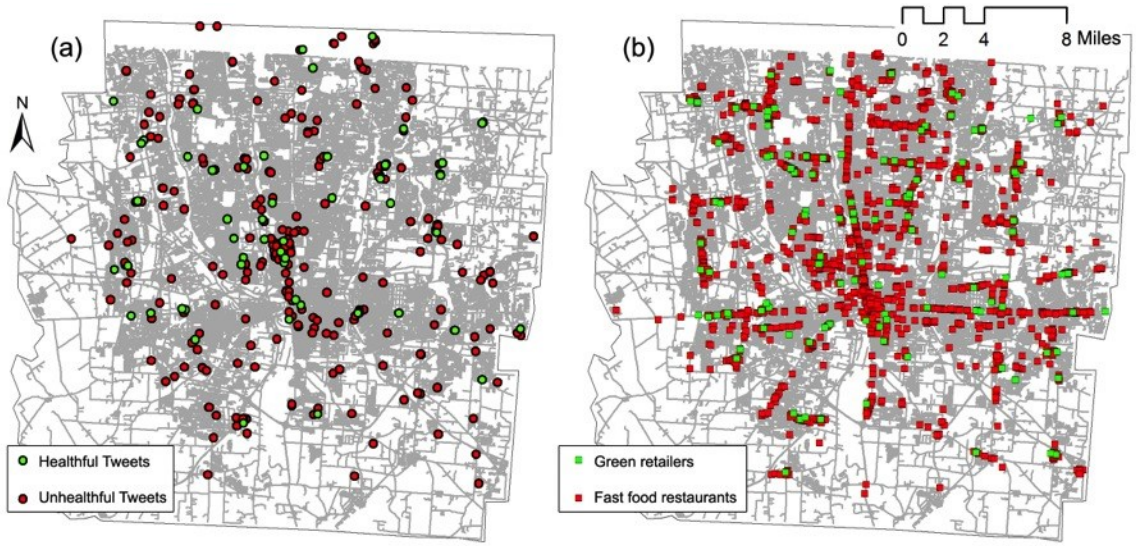


Figure 14 Distributions of selected tweets and food outlets

4.4 Results

An individual's exposure to a particular food environment can be measured in many ways, including number of nearby stores (Jago, et al., 2006; Laraia, et al., 2004), distance to the nearest store (McEntee & Agyeman, 2010), and store density, which is calculated by a kernel function (Kestens, et al., 2010). We chose to measure the quality of the food environment in terms of the numbers of green retailers and fast food restaurants within a food access zone, or specifically, within a buffered distance of the Twitter user's geo-tagged location. This measure provides a proxy of accessible healthful and unhealthful food sources within the individual's proximity. As tweets do not include individual differences in modal split, time budgets, and other socioeconomic variables that account for users' mobility for procuring food, we made use of fixed buffer distances

(0.5 miles and 1 mile) to delineate food access zones. Table 2 shows the major findings of the study in the chosen five- weekday period. In a 0.5-mile buffered access zone of reported tweets, the average number of green retailers was smaller than that of fast food restaurants. This result was due to the fact that the overall number of green retailers was smaller than that of fast food restaurants in this particular region. Between-group-comparison showed a significant difference in the average number of green retailers, 1.28 for healthful tweets and 0.54 for unhealthful tweets. This strong difference suggests that the presence of a healthier food environment may inspire a more nutritious way of eating. Interestingly, the difference in the average numbers of fast food restaurants in the two groups was not as prominent, with 19.50 for healthful tweets and 17.71 for unhealthful tweets. The slightly (and unexpectedly) larger number of fast food restaurants around healthful tweets suggests that an unhealthful food environment may not explain individuals' preference for fast food or discourage healthful food choices. Rather, the desire for fast food may be mediated by other non-spatial factors, such as financial constraints, food cues, taste of food, and lack of cooking skills that increase the motivation to patronize fast food restaurants (Driskell, et al., 2006). These results were further verified by one-way ANOVA (Analysis of Variance) tests between the subgroups of tweets: A significant difference for number of green retailers was found in healthful tweets and unhealthful tweets (p-value < 0.05), while no significant difference for the number of fast food restaurants was found in the two subgroups (p-value $= 0.46$). The sensitivity analysis was conducted by expanding the food access zones to one mile, where similar results were identified.

Another dimension of this evaluation involved examining if a change in the number of fast food restaurants/green retailers would significantly influence individual food choices. Figure 15 shows the percentages of healthful tweets versus unhealthful tweets in relation to the number of fast food restaurants (15a) and green retailers (15b) in a 0.5-mile access zone. In Figure 15(a), we could not identify an apparent pattern as to how the prevalence of fast food restaurants shapes individuals' eating habits, as suggested by Table 6. However, we found that even with low access to fast food restaurants (n < 5), a higher percentage of individuals (>85%) still tended to acquire fast food, suggesting that the choice to eat fast food is not restricted by spatial barriers but may be stimulated by personal preferences. In contrast, Figure 15(b) shows that an increase of green retailers in an individual's vicinity increased the percentage of healthy eaters at a significant rate (313.9% from n = 0 to n = 1, 8.9% from n = 1 to n = 2, 110.5% from n = 2 to n = 3) and outnumbered the percentage of fast food eaters at n = 3 (52.6% > 47.4%), indicating that spatial clustering of green retailers may help foster a healthful food environment and eventually encourage food procurement in grocery stores where more nutritious and wholesome food items are available. This observation offers corroborating evidence that the presence of green retailers has a "marked effect" on improving the quality of diets for nearby residents (Wrigley, et al., 2002).

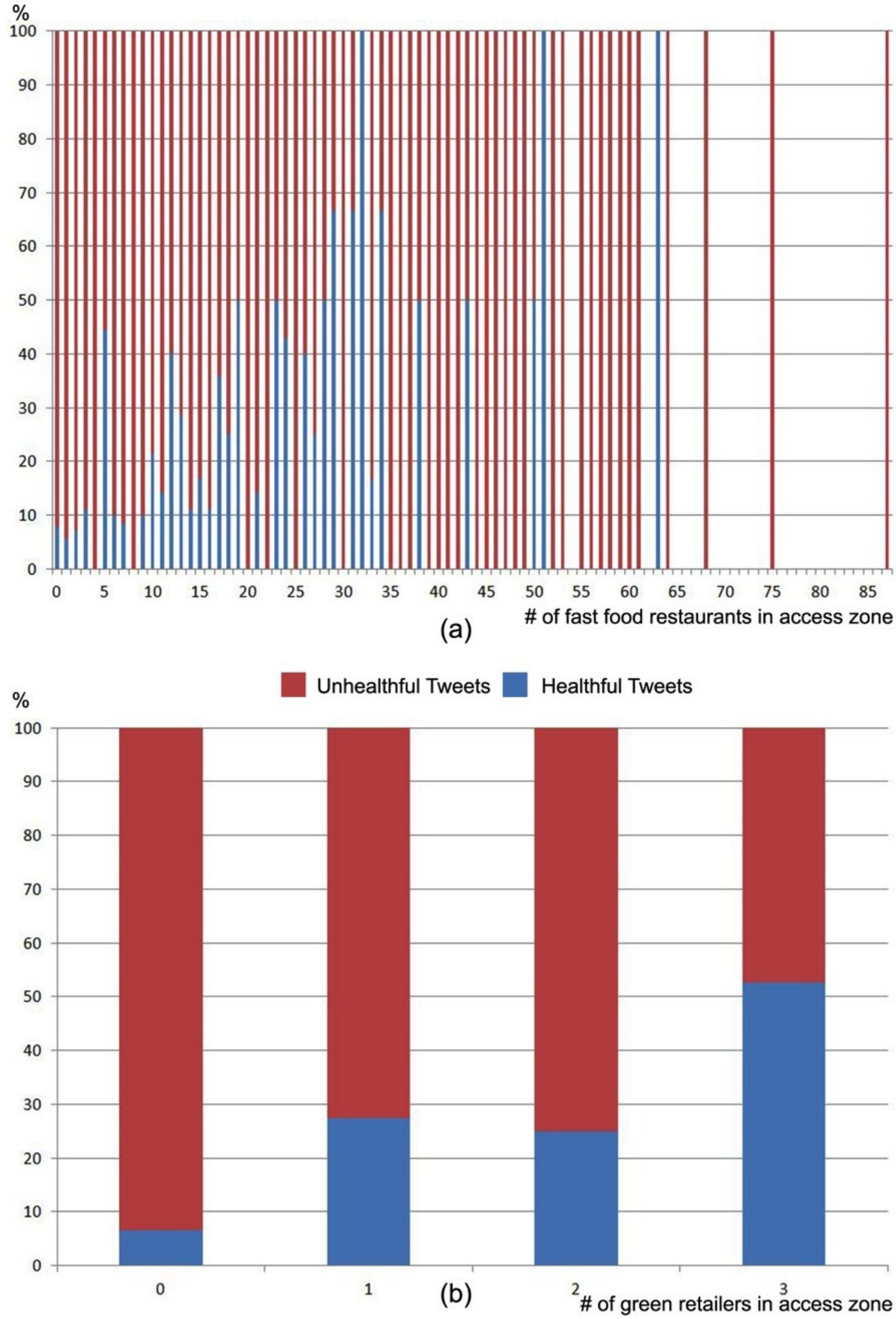


Figure 15 Percentage of healthy tweets versus unhealthy tweets

Table 6 The statistics of numbers of green retailers and fast food restaurants in a buffered food access zone between two groups of tweets

	0.5-Mile Access Zone				**1-Mile Access Zone**			
	# of Green Retailers		**# of Fast Food Restaurants**		**# of Green Retailers**		**# of Fast Food Restaurants**	
	Healthful Tweets	Unhealthful Tweets	Healthful Tweets	Unhealthful Tweets	Healthful Tweets	Unhealthful Tweets	Healthful Tweets	Unhealthful Tweets
Descriptive Statistics								
N	61	296	61	296	61	296	61	296
Minimum	0	0	0	0	0	0	0	0
Maximum	3	3	63	87	5	6	139	143
Median	1	0	18	12	2	1	39	31
Mean	1.28	0.54	19.5	17.71	2.31	1.46	44.49	41.59
ANOVA								
F	39.45		0.54		19.93		0.37	
P-value	9.82E-10		0.46		1.08E-05		0.54	
F critical	3.87		3.87		3.87		3.87	

Note: N is the number of tweets. F critical is the critical value for level α = 0.05.

These findings take on a more nuanced meaning when the issue of obesity is involved. Examining the association between the local food environment and neighborhood obesity has been recognized as having a far-reaching effect on shaping strategies for improving community health. According to several cross-sectional studies, findings regarding the link between the obesogenic food environment and obesity deviated from our perception: there was no significant relationship between density of fast food restaurants and BMI (body mass index) of the local residents (Burdette & Whitaker, 2004; Jeffery, et al., 2006; Simmons, et al., 2005), and the presence of supermarkets was found to be inversely correlated with an overweight population (Morland, et al., 2006). It is crucial to better understand the correlation between the community food environment and obesity (Holsten, 2008). This study provides a tentative explanation for this correlation, or even causality: an obesogenic food environment does not mediate unhealthy eating, which is considered a significant contributor to obesity. On the other hand, the influence of a healthful food environment is increased by expanding healthful food choices. Recognizing the significant role of green retailers is necessary for creating strategic food initiatives and land-use policies that will aid in community health awareness and will mitigate risk of obesity in the local food system. Instead of an effort to discourage consumption of fast food, public resources should be strategically allocated for improving the food environment, such as increasing the residents' accessibility to supermarkets and providing more healthful diet options that meet nutritional recommendations.

4.5 Discussion and Conclusion

4.5.1 Discussions

The emergence of social media and the infiltration of VGI have provided new channels for collecting human subject data and identifying their activity patterns. This study extracted individuals' diet choices from the widely popular social media platform, Twitter, and examined how access to the food environment around them influences their specific dietary choices. As tweets contain user-generated content and are retrievable by the public, they provide an effortless data collection process with spatiotemporally tagged information for the investigation of human activities related to food and health research. Although social media have generated a myriad of applications for public health monitoring (Chew & Eysenbach, 2010), this study is the first to focus specifically on food studies and to employ cloudsourcing of individual diet choices from cyberspace.

Social media highlights the foodscape experienced by individuals and its effects on food choices. To date, the majority of food access studies have been based on spatially generalized socioeconomic variables with very few studies investigating food exposure at a household level or around the workplace. Due to the fact that data collection poses a considerable challenge, examining individuals' food procurement and related health implications remains very rare (Kestens, et al., 2010). Soliciting food-related tweets not only overcomes barriers of traditional data collection methods but more importantly facilitates the exploration of individuals' activity space in a geographically accurate and time-sensitive manner. This possibility is further utilized by geographic approaches to measuring the relationship to individuals' perceived food environment. The conclusion drawn from this study corroborates previous findings that exposure to a healthful food environment in an individual's immediate vicinity facilitates healthful choices while showing that an obesogenic food environment may not necessarily increase the likelihood that individuals will patronize fast food restaurants. A possible underlying explanation would be that preference for fast food is primarily dictated by non-spatial mediators, such as taste and cost, as reported by a nationwide study (Glanz, et al., 1998). The real time data collection made possible by social media should be accessed to its full potential. Due to its reliance on usergenerated content, VGI on social media allows for analysis at a high spatiotemporal resolution. This availability in space and accuracy in time allows for scrutinizing human activity patterns in unprecedented detail. An example is illustrated in Figure 16, which shows the variation of healthful tweets and unhealthful tweets by time of day. It can be seen that fast food diet choices are most prevalent at mid-day until late evening from 8 PM to 12 AM, while grocery shopping is more frequent from 5 PM to 9:30 PM. This use of VGI not only demonstrates a possible method for identifying daily activity patterns by the frequency of the appearance in a given time period but also provides data for retail analysis, such as the creation of more convenient store hours. As food retailers' hours are limited, how to make complete use of VGI for balancing the tradeoff between the needs of customers and the cost of store operation would be a potential avenue for future studies.

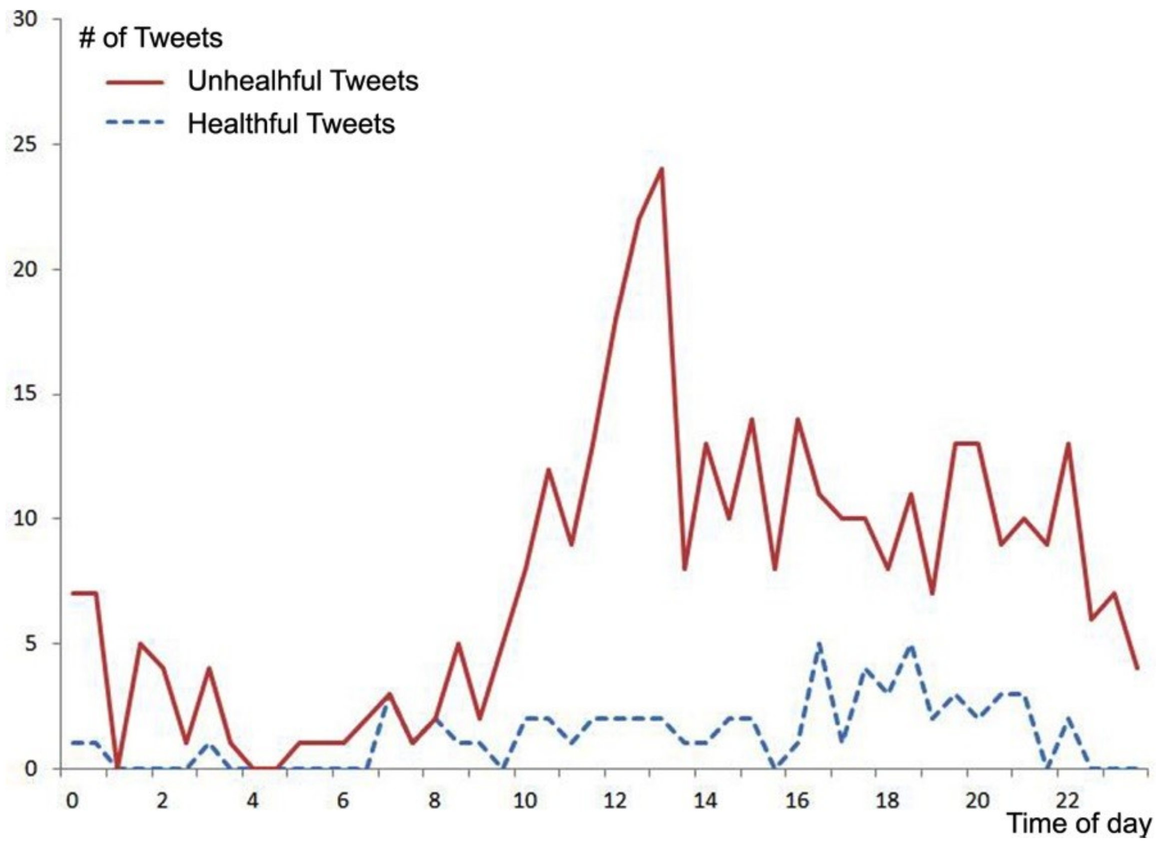


Figure 16 Variation of healthy tweets and unhealthy tweets over time

4.5.2 Limitations

Although promising finding is presented using social media (twitter) to study individual food activity, we have to acknowledge that our study also incorporates several limitations that should be taken into consideration in future research.

First, studies of human-behavior using social media data need to be held to higher methodological standards than what we have applied in this chapter. The availability of massive social media dataset as well as the powerful computational resources provides us opportunity to study human behavior at unprecedented scale. In the meanwhile, evidence suggests that the real world may be misrepresented under the analysis and forecasts using social media data. For example, the demographic composition of Twitter users is not fully representative of the entire population. Only 15% of adults who use the Internet are Twitter users, and young adults and minority groups tend to be overrepresented in tweets (Smith & Brenner, 2012). Therefore, conclusions drawn from Twitter users should also account for the compensation for variations from census data (Mocanu, et al., 2013). As Ruths and Pfeffer suggest (2014), we are fully aware the existing bias of using Twitter data in this study and we will explore methods to reducing biases and flaws in social media data in the future.

Second, the dichotomy of the diet-tweeting population (healthy vs. unhealthy) used in this chapter is just one attempt based on the authors' empirical judgment, which is not warranted in that (1) the preference for tweeting in a grocery store may not be as frequent as that in a fast food restaurant, as waiting and being seated allow for more time to tweet; (2) patronizing grocery stores does not guarantee purchase of healthful food items; (3) and fast food restaurants may offer healthful food options: examples are Panera, Piada and Chipotle. By categorizing all fast-food chains as unhealthy food has certain limitation in our study and needs further clarification. We also realize that there are some alternative ways in defining which food is healthy and which is not. For example, United States Department of Agriculture (USDA) released the federal government's new food icon, MyPlate, to serve as a reminder to help consumers make healthier food choices[10]. This website catalogues commonly eaten foods from the five food categories considered to be part of a healthy diet (fruits, vegetables, grains, protein foods, and diary). Widener and Li (2014) has applied this categorization methods in defining healthy vs. unhealthy food in their study of healthy and unhealthy food references across the United States. Studies on the actual frequency of conducting target activities should be performed to generate more convincing evidence.

Third, in urban food access studies, the maximum values for walkable distance to food retailers were loosely defined as ranging from 0.25 miles to two miles (Block & Kouba, 2006; Jeffery, et al., 2006; Mulangu & Clark, 2012), and some results were interpreted by non-spatial units, such as the percentage of travel cost in food budget (Hallett & McDermott, 2011). Based on the relatively small scale of our study, results showed that a buffer distance of less than 0.5 miles failed to capture a sufficient number of food outlets, while a distance of more than two miles showed little variation. How to standardize this cut-off value as a benchmark of sufficient food access remains to be evaluated. Also, adopting the buffer as our measurement method is to the concern of capturing the individual immediate food environment to his surrounding area. We also realize that there are other alternative ways in exploring the spatial point pattern of people's food activities such as Quadrat analysis, Kernel estimation, Nearest neighbor distance and K-function (Bailey and Gatrell, 1995).

Fourth, food choices are not only spatially constrained but also mediated by socioeconomic variables, interpersonal influences, time budgets, and individual preferences that are difficult to infer from tweets (Chen & Kwan, 2012; Glanz, et al.,

[10] http://www.choosemyplate.gov

1998; Glanz, et al., 1993; Walker, et al., 2011; Walker, et al., 2012). A possible direction to offset these uncertainties is through sentiment analysis of the contextual information in tweets (Mitchell, et al., 2013).

Last, as a large portion of tweets contain very succinct messages, compounded with typos, puns, and irrelevant content, filtering the information for the appropriate data demands a significant amount of time and labor (only 0.43% of tweets were considered relevant for our study). More efficient utilization of frontier technologies in semantic analysis to assist in data mining would be beneficial for expanding the sampling period and enlarging the sampling size.

4.5.3 Conclusion

In retrospect, building upon the conceptual trajectory of chapter 3, this chapter shifts the lens on studying a specific genre of place – food vendors under urban context with the interaction of people. Using geo-tagged tweets harvested from Twitter, with data cleaning and qualitative processing we are capable of filtering out tweets that provide perceptions of a food-related place, denoted as intersection “1” in Figure 17. The content of each tweets could be also used to extract twitter users’ daily food activity as well as their emotions when they are eating, denoted as intersection “2” in Figure 17.

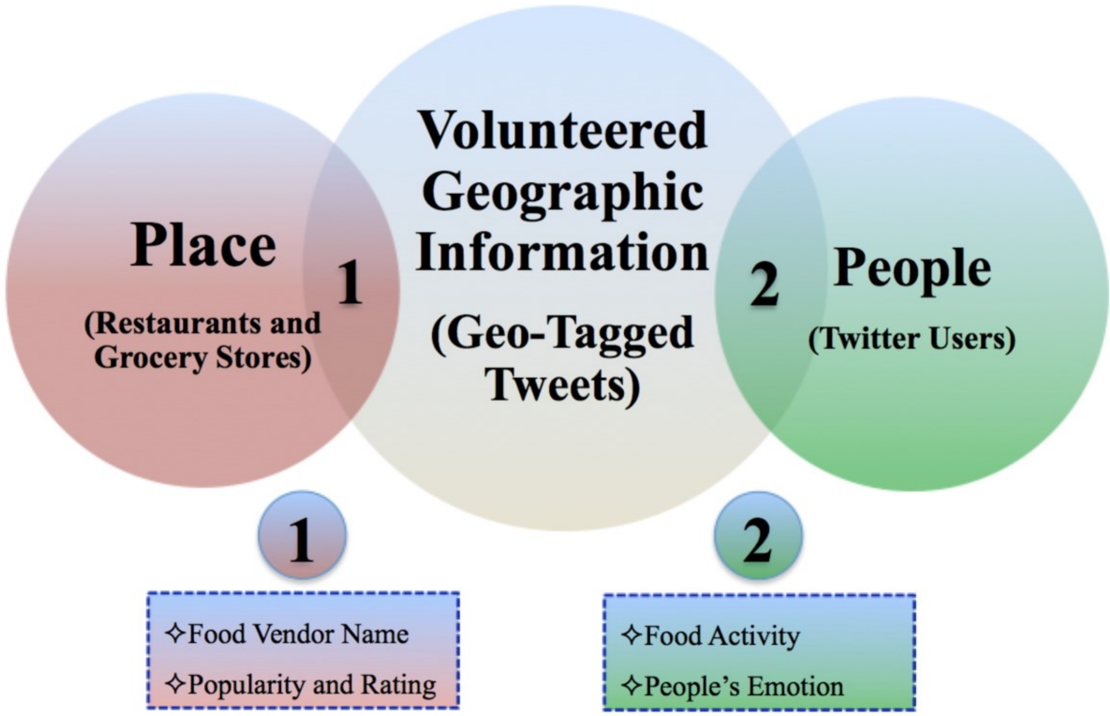


Figure 17 Framework to Study the Relationship between Food Environment and Individual Food Choice

Chapter 5: Use VGI to Study Child Beggars in Chinese Cities

In both chapter 3 and 4, we demonstrate two case studies within urban area (Columbus, Ohio) of United States using geo-social media to study the urban pulse in general and people's food activity in specific. In this chapter, we shift our lens to investigate a VGI effort to study the social inequality issue of children in China. In particular, using a representative sample of microblog posts from Chinese mainstream social media as well as contextual information from a variety of sources, this chapter examines the geographic distribution of street children in China as well as identifies how VGI are capable of empowering citizens to help those children find their way back home.

5.1 Introduction

Being the world's most populous country in the world with growing disparities between regions and various social groups, child begging are quite common in China and has become a major social issue with growing national attention. A report by the National Working Committee on Children and Women indicates that China has approximately 1 to 1.5 million homeless children[11]. Among these homeless children, China estimates 20,000 children are victims of abduction every year[12]. Even worse, some are forced to beg by criminal organizations, while others are sold to couples who cannot have children, or are made to work in factories as child labor.

---

[11] http://cnzgw.org/2012/0629/8761.html

[12] http://www.foreignpolicy.com/articles/2011/10/06/china_missing_children

The Chinese government has been actively pursuing programs to effectively address the child beggar issue. In 2003, the central government issued a national policy on the *Management of Street Beggars in Cities*, after which assistance stations were established to provide food and shelter for the homeless in major cities. The government has launched five major crackdowns on human trafficking since the 1990s. In 2010, a national DNA database was established to assist the identification and rescue of child beggars. Blood samples from parents of children reported missing and samples from children of unclear background were collected and stored for automatic matching. These actions initiated by the government have had very positive impacts on the child beggar issue. According to the Ministry of Public Security's report, more than 9,300 kidnapped children in China have been rescued since April 2009 after the nationwide campaign was launched to crack down on human trafficking[13].

However, the Ministry noted in the report that children kidnapped to become beggars took up only a small portion of all cases of child beggars. In a growing number of cases, children were taken to beg along with their parents or relatives. As a result, the government encouraged the involvement of citizens in providing clues to help the police rescue children - especially those being abused and forced to beg on the streets.

Street children have gained attention of scholars as a research topic, and yet traditional studies of street children have certain limitations. First, these studies are often based upon very small number of samples that can hardly provide adequate representation of the entire population of street children. Also, previous studies of street

[13] http://www.legaldaily.com.cn/index_article/content/2011-02/11/content_2467637.htm?node=5954

children focus mainly on developing theories and conceptual frameworks, with less emphasis on empirical studies. Furthermore, traditional studies rely primarily on using interviews and surveys to collect the data of street children, which always face issues such as privacy and unwillingness to report.

Different from traditional approaches that collect data through individual surveys and interviews of studying the street child reported in the literature so far (Bromley and Mackie, 2009; Panter-Brick, 2002; Cheng and Lam, 2010), we use a representative sample of microblog posts from Chinese mainstream social media as well as contextual information from a variety of sources. We will examine the geographic distribution of street children in China as well as identifies how VGI are capable of empowering citizens to help those children find their way back home.

5.2 Related Work

Until recently virtually all geographic information was produced through a top-down process by government mapping agencies or corporations in the mapping industry. The geospatial technologies that emerged since the 1960s did little to change this set of arrangements, since their major impacts were on the acquisition of raw data through new and more efficient instruments. The transition from paper-based to digital dissemination, from paper map sheets to tapes and eventually Internet distribution, left most of the arrangements intact (Goodchild and Glennon, 2010). By the mid-1990s, however, new technologies were emerging that had fundamentally changed these arrangements. First, it has become possible for the average citizen to determine positions accurately these days, without formal training. Second, it has also become possible for almost anyone who has access to the Internet to develop the skills to make maps from acquired data - cartographic design skills previously possessed only by trained cartographers. Google's Maps service, for example, allows anyone to produce a decent-looking map from custom data, and OpenStreetMap render raw data provided by the user into a cartographically acceptable street map. Increasing, citizens with no expertise in the mapping sciences were suddenly able to perform many of the complex mapping tasks that had previously been the preserve of experts. Volunteered geographic information (VGI) (Goodchild, 2007), also known as neogeography (Turner, 2006) or neocartography (Liu and Palen, 2010), has been coined to describe this phenomenon. The world of volunteered geography is breaking down the traditional distinctions between expert and non-expert in the specific context of the creating geographic information, since many of the traditional forms of mapping expertise can now be performed through various on-line or hand-held devices.

Many web sites have emerged in the past few years to encourage and facilitate the actions of neogeographers. In essence these sites make it possible for user generated content on the Web to become useful and usable geographic information for various mapping applications. Popular sites include Flickr and its georeferenced photographs, the OpenStreetMap, Wikimapia and its large collection of user-described features, and numerous citizen science web sites that collect georeferenced observations on plant, animal, and bird sightings. Furthermore, it is increasingly common for the content of Twitter, Facebook, and many other social media to be georeferenced. With the social media websites increasingly location-based, most location-based social media currently allow users to create and share geo-tagged information about their real lives and we now have deep data about many for social and behavioral study as discussed by Sui and Goodchild (2011). Using data harvested from location-based social media represents a major advances in methodologies of social science research (Gupta et al, 2012; Tsou, 2012).

In the context of this study, we found the following sites are of particular relevance:

- Street Photos to Rescue Child Beggars.

Child begging phenomenon is drawing national attention due to the growing popularity of social media among the Internet users in China. Photos posted on Sina WeiBo (microblog) raised the level of public awareness on the child beggar issue in China since winter 2011. The child beggar issue first gained attention by a request from a mother who lost her child in 2009 and sought help from the public using social media. Based upon the clues of several photos posted on Sina WeiBo, the lost child was found begging in the central business district in Xia'men. Prof. Yu Jian-Rong from the Institute of Rural Development, Chines Academy of Social Sciences (CASS), saw this piece of news and was motivated by the huge potential of crowdsourcing efforts in helping the parents to find their missing children. On January 25th, 2011, Prof. Yu created a public blog on Sina Weibo called "Street Photos to Rescue Child Beggars"[14] to encourage the citizens to take photos when noticing suspicious kidnapped child begging on the street and spread these photos through its public site on Sina Weibo. The blog received large amount of volunteers' responses in a short period, more than 200,000 followers and 7000 posts by August 31st, 2012.

[14] http://www.weibo.com/jiejiuqier

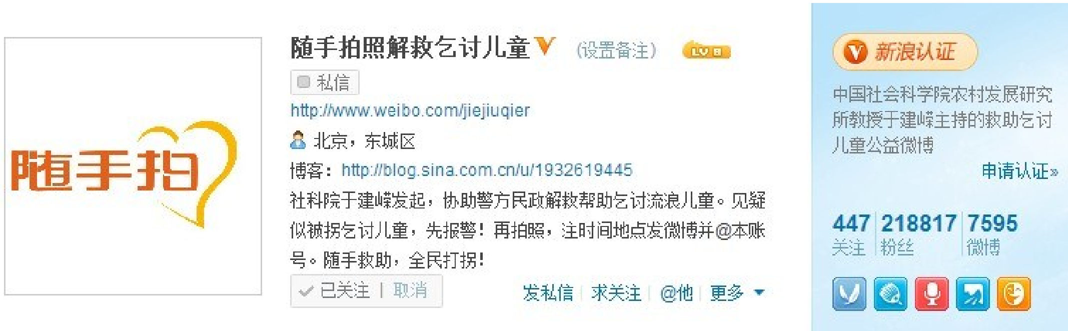


Figure 18 Weibo "Street Photos to Rescue Child Beggars"

With the availability of this relevant, timely and loosely-coupled data from social media, we are now capable of exploring this social problem from a geographic perspective. Since the child beggar problem has been reported across the country, our first question is how these child beggars observed on the street are spatially distributed around the nation? Our hypothesis is that the child beggars are not evenly distributed, and there exist distinctive patterns of geographic distribution. Our second objective is to examine what socio-economic factors can be used to explain the distribution of the child beggars reported by the social media.

- Baby Back Home.

China's Ministry of Public Security has set up a national DNA database, connecting 236 DNA centers nation-wide, to help track children abducted by human traffickers. The idea of a database came from Zhang Baoyan who heads up a website called "BaoBeiHuiJia" (http://www.baobeihuijia.com), whose focus is the return of abducted children. The Chinese government investigates 3,000 missing children cases a year, although the actual figure may be as high as 20,000 children abducted annually, said Xinhua quoting a US Department of State statistic[15]. Many children are sold within China as extra labour, household servants, workers in mines and brick kilns, or into illegal prostitution rings. Baobeihujia.com has registered about 10,000 users who have put information about missing children online. It has successfully helped more than 500 children to find their parents since it began in 2007. *Baobeihuijia* means "baby back home," which contains large amount of qualitative/narrative data that can be mined and analyzed for this chapter.

Figure 19 Website of the "baobeihuijia"

5.3 Methodology

This dissertation relies on a combination of quantitative as well as qualitative data from social media, supplemented by socio-economic data from government sources. We also used a combination of quantitative and qualitative analysis methods to examine the spatial and temporal distribution and mobility patterns of Child beggars in China.

- Volunteered Geographic Information.

[15] http://www.foreignpolicy.com/articles/2011/10/06/china_missing_children

We have two sources of volunteered geographic information and we develop two corresponding strategies to collect the data used in this chapter. For the "Street Photos to Rescue Child Beggars" website, each post with a street child beggar is acquired. Similar to Twitter, Weibo has a public available API. To take full advantage of the new geographic data directly generated by the volunteers on the web, we prepare a data process mechanism in order to convert the data to the conventional format that can be used for geographic analysis. The raw data is usually a paragraph of text with the location at which the photo was taken and a descriptive paragraph of the observation. Some data are also found with attached one or two photos. These raw data were collected and interpreted via a java application written using Sina Weibo API[16]. We built a crawler to retrieve all the post and we name this set of data as dataset 1. For the website "BaoBeiHuiJia", we used its built-in search engine to search the website to retrieve all the reported missing children. Records are retrieved into the database with the attribute of "ID", "name", "sex", "birthday", "native place", "missing place", "missing time" and description. We name this set of data as dataset 2.

- Socio-economic Data

Our study of the child beggars is set to be focus at the provincial level. As a result, a 2006 Chinese administrative boundary map in shapefile format is acquired from National Fundamental Geographic Information System of China[17] for our study. All the child beggar observations were geo-coded and aggregated to the provincial level for further analysis. In order to examine the possible socio-economic factors that influence

[16] http://open.weibo.com/wiki/API/en
[17] http://nfgis.nsdi.gov.cn/nfgis/english/default.htm for details

the distribution of the child beggars observed, we collected four potential variables for each province respectively. A joint function was performed using ArcGIS to relate these variables to the boundary data for performing the exploratory spatial analysis.

Table 7 Socio-economic variables and distribution of child beggar

ID	Province	Area (Sq Mile)	Weibo_CT	BBHJ_CT	Pop_Density	HDI	Internet_P	GDP
0	Beijing	6384.08	204	639	0.11358	0.891	60.0	82472203
1	Tianjin	4347.51	37	77	0.19887	0.897	43.5	58463896
2	Hebei	72079.46	54	667	0.00503	0.810	19.2	165346610
3	Shanxi	60284.69	26	394	0.00339	0.775	24.1	64693502
4	Neimenggu	443520.03	40	162	0.00005	0.803	16.0	76594970
5	Liaoning	56656.08	59	193	0.00526	0.835	26.5	152596879
6	Jilin	73910.31	4	194	0.00193	0.776	19.0	64252240
7	Heilongjiang	173834.03	19	256	0.00050	0.786	16.2	84784599
8	Shanghai	2249.30	247	221	1.12827	0.908	59.7	116781919
9	Jiangsu	38442.57	233	917	0.01941	0.805	27.3	329702200
10	Zhejiang	38505.62	101	430	0.01196	0.841	41.7	193878475
11	Anhui	54145.16	97	551	0.00820	0.727	11.8	90177425
12	Fujian	46718.84	246	811	0.00589	0.801	38.5	111106664
13	Jiangxi	64814.30	45	488	0.00384	0.735	14.0	91933468
14	Shandong	59994.23	85	858	0.01003	0.828	21.2	335665342
15	Henan	64113.88	97	1328	0.00890	0.787	13.7	185655039

Continued

Table 7 Continued

ID	Province	Area (Sq Mile)	Weibo_CT	BBHJ_CT	Pop_Density	HDI	Internet_P	GDP
16	Hubei	71855.26	125	646	0.00447	0.755	18.4	134160512
17	Hunan	81604.74	237	509	0.00390	0.751	15.7	117774357
18	Guangdong	68154.91	774	1001	0.00579	0.828	48.2	261699158
19	Guangxi	91514.70	73	405	0.00216	0.741	15.4	70061593
20	Hainan	13166.73	97	37	0.01618	0.762	25.6	12480298
21	Chongqing	31815.96	136	600	0.01151	0.675	21.2	69146882
22	Sichuan	217542.02	172	1517	0.00097	0.728	13.6	246712703
23	Guizhou	68065.30	97	1318	0.00299	0.647	11.5	31672460
24	Yunnan	148249.58	71	365	0.00072	0.672	12.1	51569518
25	Xizang	438126.19	4	3	0.00000	0.621	16.4	3372973
26	Shaanxi	78577.56	54	511	0.00228	0.775	21.1	96976067
27	Gansu	146829.66	22	147	0.00044	0.681	12.5	30020860
28	Qinghai	283435.53	42	51	0.00002	0.684	23.6	8588187
29	Ningxia	20041.91	8	71	0.00496	0.724	16.6	9230209
30	Xinjiang	629224.88	19	96	0.00002	0.757	27.1	33409418
31	Taiwan	14058.33	4	0	0.04424	0.932	70.0	242708378
32	Hongkong	400.50	4	0	13.97410	0.862	70.0	167374865
33	Macau	1.73	0	0	50136.30000	0.940	66.0	7244222

Where:

GDP: Gross Domestic Product in 2009 of the corresponding province

Population: Total population reported in 2006 from National Bureau of Statistics of China

HDI: Human Development Index reported in 2009 from China Development Research Foundation.

Internet_P: Internet Penetration Rate reported in 2009 from China Internet Network Information Center.

- Exploratory Spatial Data Analysis.

Exploratory Spatial Data Analysis (ESDA) is a set of techniques aimed at describing and visualizing spatial distributions, at identifying a typical localizations or spatial outliers, at detecting patterns of spatial association, clusters or hot spots, and at suggesting spatial regimes or other forms of spatial heterogeneity (Haining 1990; Bailey and Gatrell 1995; Anselin 1998a, 1998b). These methods provide measures of global and local spatial autocorrelation, which can reveal spatial clustering patterns.

To answer our research questions, we first performed the Exploratory Spatial Data Analysis (Anselin, 1998) trying to find out the global trend and pattern of the distribution of child beggars observations count for each province. To test whether there is spatial dependence existing for the count of child beggars in each province, Moran's I was then calculated toward the count using the definition:

$$I=\frac{N}{\sum i \sum j \omega_{ij}} \frac{\sum i \sum j \omega_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum i (X_i - \bar{X})^2}$$

Where N is the number of total provinces (34) indexed by i and j. w_{ij} is an element of a matrix of spatial weights and we generated using a conventional Queen's method. A stepwise regression was run to examine the possible contributing factors that may explain the distribution of child beggars.

- Spatial-time paths approach.

Time geography was first proposed by Torsten Hägerstrand in the 1960s to study human activities and their constraints under the context of space and time. The space-time system is represented as a three-dimensional (3D) system which consists of two spatial dimensions and another temporal dimension. A space-time path, which is defined as a trajectory that records an individual's movements in space and over time, is a critical concept in time geography. We have already seen significant progress of implementing the time geography concepts in a GIS environment over years (Kwan and Hong 1998). For a better visualization purpose, a generalized spatial-time path tool was developed by Shih-Lung Shaw and his team (Shaw, Yu and Bombom, 2008). The space-time path approach provides an efficient method of organizing individual mobility histories in an integrated space-time environment. In this chapter, we used the space-time path tool developed by Shaw to visualize the mobility patterns of child beggars[18].

5.4 Results

- Geographic Distribution.

The spatial distribution of child beggars is mapped using ESRI's ArcGIS. To examine the spatial distribution of child beggars and missing children, we created a bivariate map of child beggars and missing children with population density as the

[18] http://web.utk.edu/~sshaw/NSF-Project-Website/download.htm

backdrop (Figure 20 and Figure 21). The results using data from both Weibo data and BBHJ data show that the child beggars and missing children observed and posted on the web are mostly concentrating in the south-eastern part of China, indicating that either child beggar cases exist more in these areas or more people from these areas are willing to voluntarily contribute. In contrast, the west China (Xinjiang, Xizang, Gansu and so on) receives limited count of the child begging phenomenon. Also, provinces with high population density tend to have more child beggars and vice versa.

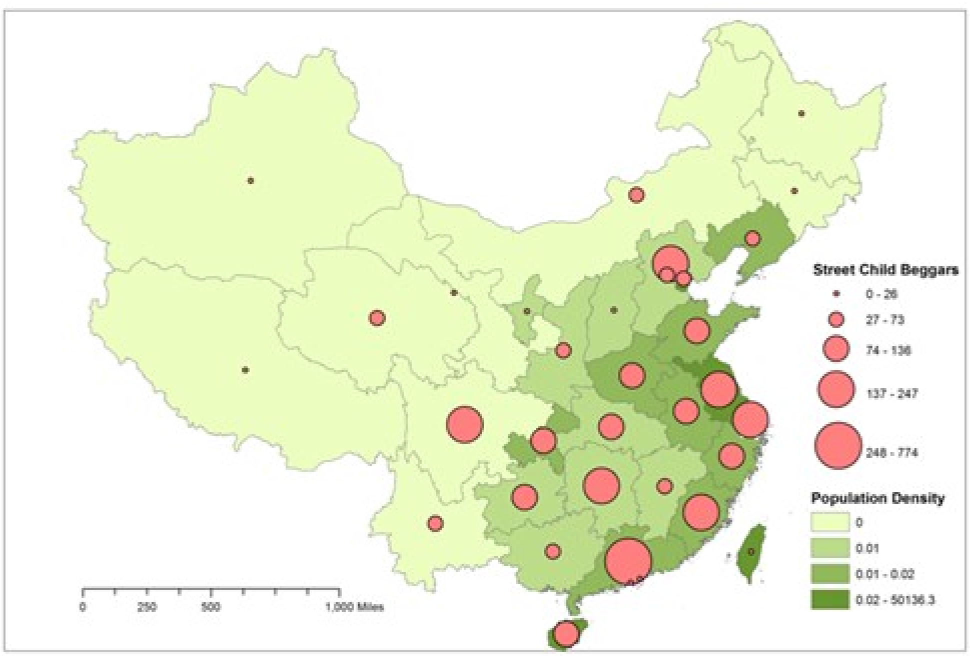


Figure 20 Spatial Distribution of Child Beggars in China, 2011

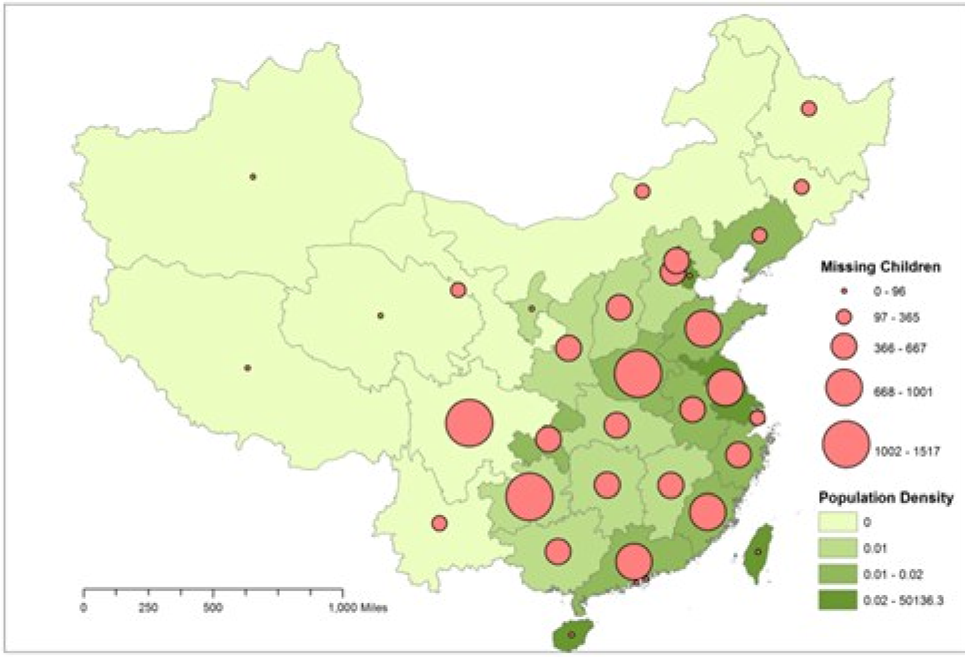
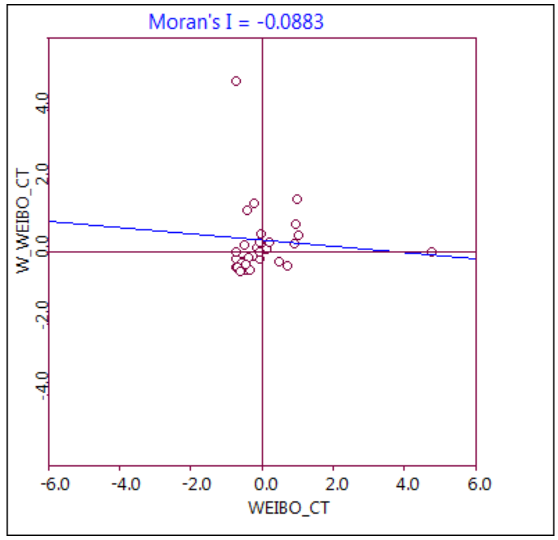


Figure 21 Spatial Distribution of Missing Children in China, 2011

- Exploratory Statistical Analysis.

Before performing the regression, we need to determine what's a better regression model we would use by looking whether spatial dependence exist of the count of Weibo and count of Baobeihuijia in each province. The univariate global Moran's I (Figure 22) shows that there is no spatial autocorrelation so we decide to use the ordinary least square (OLS) to test the regression model.



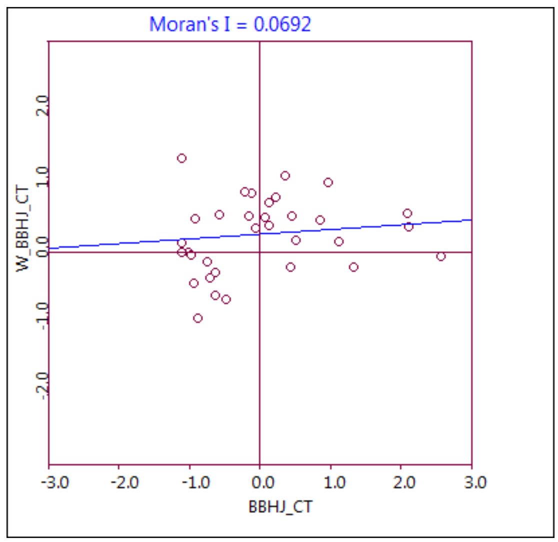


Figure 22 The Moran's I show weak spatial autocorrelation

We perform a stepwise regression to study the relationship between the number of child beggars/missing children and several socio-economic variables. The number of child beggars and missing children aggregated at province level is set to be the dependent variable and Gross Domestic Product (GDP), Human Development Index (HDI), Population Density (PD) and Internet Penetration Rate (IPR) are four independent variables. The result in Table 2 indicate that HDI is positively correlated with the number of child beggars in each province.

Table 8 Results of step-wise regression

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-76.0989	1059.728	-0.07181	0.943244
HDI	995.3628	1547.934	0.643027	0.525259
POP_DENSIT	-0.00393	0.009494	-0.41438	0.681644
INTERNET_P	-6.81648	8.838534	-0.77122	0.446814
GDP	-0.00114	0.002259	-0.50437	0.617811

- Spatial Mobility Trends

Human mobility is the physical movement by humans from one area to another. Using the successful cases from dataset 2, we are able to depict the lost and found link of each child (Figure 23). According to the map, we are able to confirm the home places in most cases of missing children are located in Western and Central China while these children are often found in Southern and Eastern China. This finding indicates a general mobility trend from the less developed regions of China in the West and North to well developed areas in the East and South.

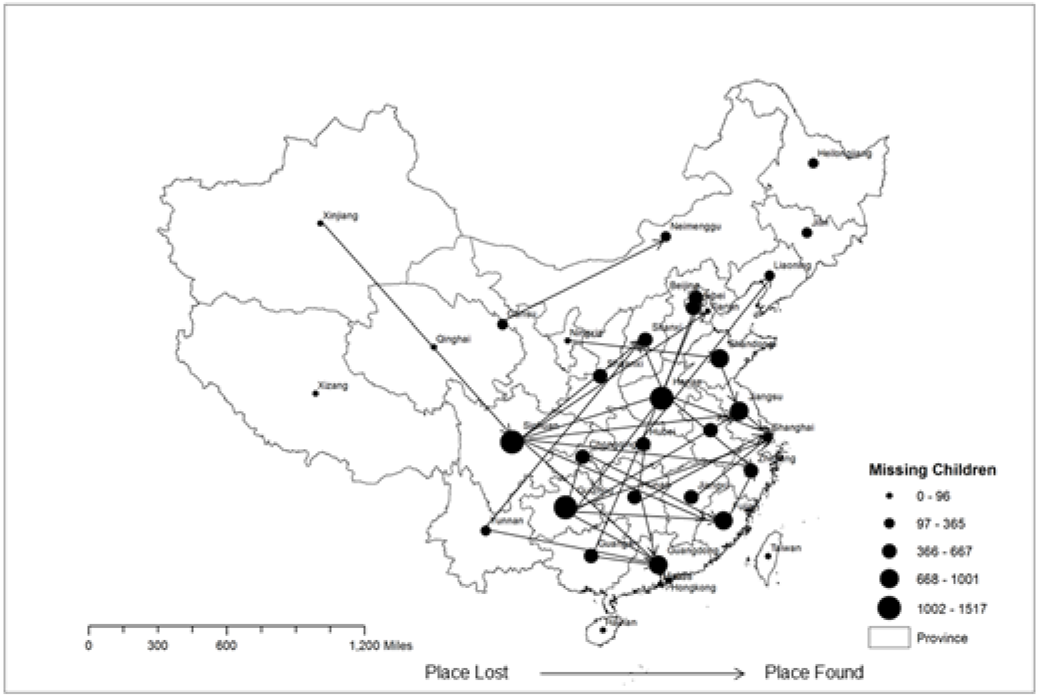


Figure 23 Lost-Found places of missing children in China

- Temporal Distribution Patterns

In order to analyse the social impacts of the “Street Photos to Rescue Child Beggars” activity, we collected the total number of postings in each month since the launch date (Jan 25 2011). As we can see from Figure 24, the amount of cases posted had been relatively steady throughout the year – 100 to 400 postings in most months except February. The number of cases soured to 3464 entries in February, which marks the beginning of the Chinese Lunar calendar year and the Spring Festival. Apparently, this can be explained along multiple lines. It is quite possible that during the month of the Chinese new year - the most important holiday in China (equivalent to Christmas in the

West), most people are keen in having family reunion, thus becoming more sensitive and earnest in finding their own ways of addressing the child beggar and missing children issue. Another possible reason is that most of the people in China during the month of Chinese new year are on break, thus they have the luxury of time to go out to take the photos, blog/post them, and respond to calls of actions from various social media sites.

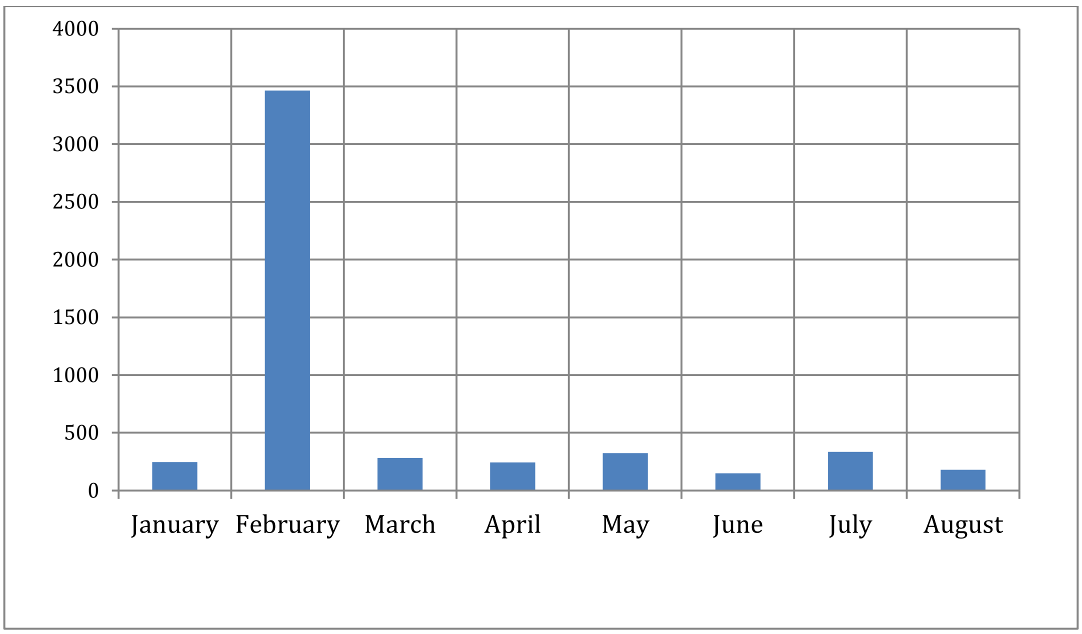
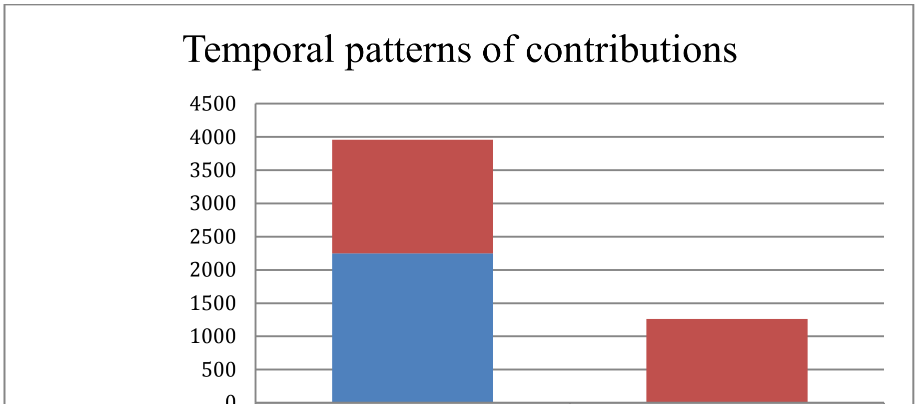


Figure 24 The amount of posting by month

Also, we conducted a temporal analysis to capture the behavioural patterns from the contributors. The count of the contributions received in weekdays versus the ones received in weekends is summarized (Figure 25). Also, we compare the amount of contributions from the working hours to the ones from the non-working hours as well. Apparently, three-quarters of the postings were made during the weekdays and only about a quarter of postings were made during the weekends. It is also interesting to note that postings made during working hours are equivalent to those made during weekends plus non-working hours during the weekdays.



	Weekday	Weekend
Non-Working Hour	1710	1263
Working Hour	2249	0

Figure 25 Temporal Distributions of VGI Postings

"Street Photos to Rescue Child Beggars" gets its attention from different cities in different time. Besides the unevenness of the spatial distribution around the country, there is also variation in terms of the timing of microblog postings from different cities. For example, the first post was from Beijing in January 25th, 2011. Then a volunteer from Wenzhou contributed a photo in the same day. Within 10 days, this social media website received posts from 71 different cities in China. In order to visualize the space-time trend of the VGI distribution, 71 county-level cities are identified sequentially based upon the time of its first contribution posed. To simplify their timestamp information, we denote the sequential time using a relative time stamp T1, T2 ..., and T71. Along with their locational information, a space-time path has been mapped using Shaw's Extended Time-Geographic Framework Tools Extension in ESRI Arcgis 9.3.[19] (Figure 26). Although the rhythms of VGI contributions may not be so obvious in our case due to the relative short time span, the space-time path approach nonetheless demonstrates its great potential as a tool to visualize the spatial-temporal patterns of VGI contributions via social media sites.

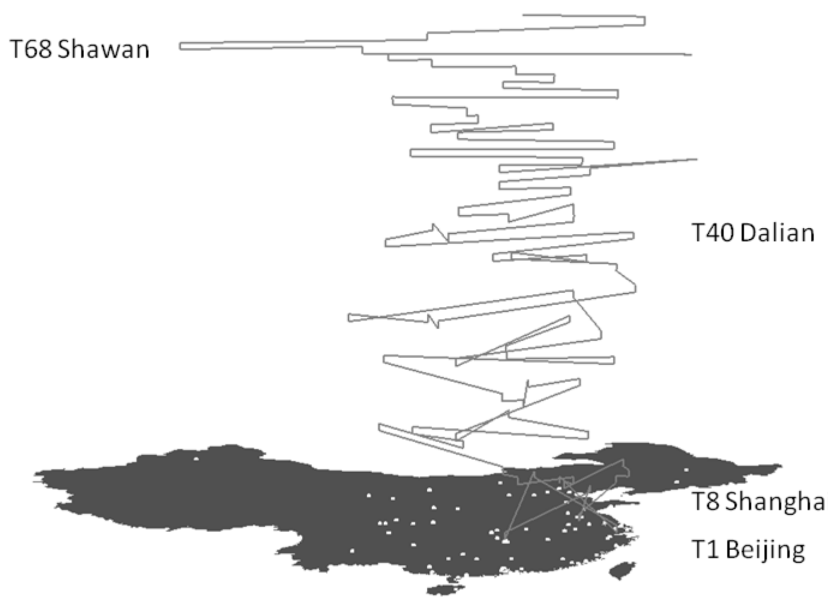


Figure 26 The space-time path of VGI contribution on social media sites

- Narrative Analysis.

[19] http://web.utk.edu/~sshaw/NSF-Project-Website/download.htm

As a useful qualitative approach, narrative analysis or inquiry is “the study of experience as story” (Cornnelly and Clandinin, 2006). This research method could not only capture people’s experience as stories, but also reflect the social, culture and institutional contexts where the experiences were constituted (Clandinin and Rosiek, 2007). Kwan and Ding (2008) bring out a geo-narrative approach, which integrates qualitative geographic information system and narrative analysis.

Our study finds out the potential to apply the narrative analysis approach on successful cases from “BaoBeiHuiJia” website. Taking the description from missing children who seek help to find out their biological parents as narrative material, we can not only capture the geographical information from the text, but also infer the social and cultural context of the place and circumstances that triggered the cases of missing children. Figure 27 shows an example of narrative analysis of a child “San Wa” who was abducted in 1996 from his hometown. Four different colours are used to mark the temporal reference, locational reference, action and feelings respectively. In this example, the child was abducted on Feb 22nd, 1996 when he was in his age of five. He didn’t remember the exact name of his hometown, but his memory recalled the name is “Wei” county. He was first abducted by a man with a beard to a valley, and then he was sent to his foster parents place, Xing Tai city in Hebei province. He vaguely remembered the name of his father and brother. Also he pointed out some of the feature of his hometown such as there is a white elephant statue over a school. With the narrative information, volunteers from the web community successfully helped “San Wa” find his biological parent in Yunnan province.

1996年2月22日，五岁多的三娃在家门口小河边，一名大胡子男人和那个小女孩母亲一同将三娃带走。大胡子人贩先将三娃带到一个山沟。买家因近临盆，与人贩商议，如生女，便留下三娃，如生男，便让人贩带着三娃另寻买家。三日后，买家生下一名男婴，人贩领着三娃，辗转一天一夜火车，续又坐客车，将三娃卖于现养父母家——河北省邢台市任县。三娃儿时记忆未减，他记得自己的家乡好象叫威县？渭县？蔚县？不能确定是那里的。爸爸叫陈泽明或陈泽民。有一个哥哥一个姐姐。哥姐大自己四五岁，上学年纪。哥哥叫陈什么龙。自己排行老三。他还记得他有爷爷奶奶。一个祖婆，奶奶叫家家，爷爷叫家公。天冷时，奶奶就烤红薯给三娃吃。儿时记忆，夜晚萤火虫点着亮光低飞，一家人就围着炉火吃饭，木桶蒸饭的味道三娃至今还记得。奶奶还常用竹筒捣辣椒。他常和哥哥姐姐去县城街边玩，哥哥姐姐的学校就在那附近。那里还有一个石碑色长牙大象，每次玩乐，都很尽兴。

Figure 27 Narrative analysis from description material of “San Wa”

5.5 Discussions

Street children, usually referred as kids who permanently live on street to the extent of even sleeping there at night, are often excluded from the mainstream society and off the radar screen of government census. These particular groups of children are often out of school and do not have access to basic health services or the protection of an adult and they work, beg and steal for a living on the street (De Venanzi 2003). Because the children living on the street has drew emotive public concern and media coverage, the study of street children has become a matter of priority for national and international child welfare organizations and thus receive attention by academic researchers (Panter-Brick 2002). For example, Bromley and Mackie (2009) explored the experiences of over a hundred child traders on the street in Peru, suggesting that international policy changes should be made in order to better protect the street child worker. Young and Barrett (2000) studied the interactions between street children and their socio-spatial environment by using four visual methods which include the mental and depot maps, thematic and non-thematic drawings, daily time lines and photo diaries. Sociologists find out that street children tend to have a lower level of subjective wellbeing than those regular children after an examination of street children in China (Cheng and lam 2010).

Studies on children and childhood have become significant research topics for social scientists in general and geography in particular (Adebe 2009). In geography, a social study about the future of children had been conducted in Bunge's (1971) classic "Fitzgerald: The Geography of a Revolution" some 40 years ago. Hart (1978) conducted a naturalistic and descriptive study towards a small group of children about their life experiences and shared some common characteristics in both space and place. To better understand the lives of younger people under the age of 25 from a geographic perspective, an interdisciplinary journal was launched in 2003, devoted to discussing issues that have significant impacts on children's lives. There exists different ways in which children can be involved in the research process (Adebe 2009). Generally speaking, the different methods of conducting children's research were identified by Chirstensen and James (2000). According to their classification based on the participation levels, children may be treated as objects, subjects and participants respectively. This study is a major extension of studies on children's geography by deploying data from social media, coupled with both quantitative (GIS-based spatial analysis) and qualitative approaches (narrative analysis). Despite the preliminary nature of our results, we obviously can detect multiple stories behind these results, which may point to directions for future studies in this area.

- Growing economic and regional disparities in China.

As the results of our analysis indicate, the overwhelming majority of child beggar cases reported by volunteers in Chinese social media (e.g. Sina Weibo) take place in cities in the eastern party of China, but the children themselves are either from rural areas or Western provinces in China.. According the "three economic belts" (Figure 28) scheme based on the Seventh Five-Year Plan (1986-1990), east China tends to be economically more developed than west China and the gap between urban and rural areas in China also tend to enlarging rather than narrowing in recent years Since a common hypothesis is that street child beggars would prefer a location with higher level of economic development, the spatial distribution of the child beggars in fact reveal the growing regional and urban-rural disparities in China. Despite improvement in economic development levels across China in recent years (Fan and Sun, 2008), Gini coefficient in China – one of the leading indicators for regional and social disparity – reached a dangerous level. According to the World Bank (2005), the Gini coefficient, has risen from 0.33 in 1980 to 0.47 in 2005. Many other studies have also provided strong evidence on the growing regional inequality in China (Wei and Ma, 1996; Fan and Sun, 2008) using the social-economic statistics data. The geographical distribution of child beggars is a manifestation of this growing pain in China – an important issue China must confront in the coming years in order to achieve the goals of sustainable development it set for itself.

- VGI, Social Media, and People-based GIS

Methodologically, this chapter aims to cross the quantitative and qualitative chasm by using volunteered geographic information (VGI) harvested from social media sites. Through a combination of GIS-based quantitative spatial analysis and narrative-based qualitative approach, this chapter revealed multiple dimensions of the child beggar issue in China that would be almost impossible due to the lack of needed data. We are fully aware that there are multiple challenges that need to be addressed in using VGI for geographic research before its full potential can be realized (Elwood et al., 2012). This chapter is a modest step toward that direction. Furthermore, as more and more detailed data are available at the individual level tagged with explicit spatial as well as temporal information, we are apparently moving closer to what Miller (2003) envisioned as the "people-based GIS." With location-awareness technologies increasingly embedded in mobile phones, cameras, pedometers, and other hand-held devices, citizens can nowadays record their daily activity with accurate location in real time and instantaneously shared with the rest of the world. These massive individual level data with spatial and temporal information is rapidly merging with various other streams of big data and will serve as a gold mine for GIScience researchers to practice people-based GIS. In addition, with the ever expanding blogosphere and various social media, we will be accumulating a massive amount of "deep" qualitative data and stories about people's lives as well (Manovich, 2011). The emergence and growing popularity of social media could give us the chance to have deep data for many (Sui and Goodchild, 2011) to practice the so-called people-based GIS. The fusion of GIS and social media could help us not only map the distribution of spatial phenomena, such as the uneven distribution of child beggars in China, but also trace the relevant trajectories of certain people or events. For example, when a child beggar is posted on social media sites, concerned citizens will help to spread the word so that they can gain more public attention. Sometimes, people may also add comments or evidences to the original post which may help to rescue the children. Successful cases have been reported through this thematic account that some begging children are recognized by their biological parents later or rescued by the local police.

Figure 28 Three Economic Zones in China

5.6 Summary and Conclusion

The goal of this dissertation was to explore the spatial and temporal mobility patterns of street child beggars in China using volunteered geographic information as our primary data source. Volunteered geographic information harvested from the leading

Chinese social media website – Sina WeiBo (Microblog) were used for our empirical analysis. Exploratory spatial data analysis (ESDA) was used to explore the spatial patterns of child beggars in China. Because no spatial autocorrelation was detected using univariate Moran's I, we implemented an ordinary least squares (OLS) linear regression model to find which socio-economic variables would be used to fit the model. The results show that child beggars are mostly observed in cities in South-east China – the most prosperous and well developed regions in China. Development levels as measured by HDI is found to be two critical factors that that explain the spatial distribution of child beggars in China. The study also finds out VGI varies across time and a space-time path is built to facilitate our hypothesis. Our preliminary results have revealed interesting spatial and temporal distribution and mobility patterns for street child beggars in China. The geographical patterns are consistent with the growing regional and social disparities in China. By engaging in more detailed spatial and temporal data at the individual level, we are moving a step closer to people-based GIS.

There are limitations for this project and further studies are warranted. First of all, the volunteered geographic information are collected and aggregated at the provincial level, which served our goals well because we want to understand the macro spatial and temporal mobility patterns of child beggars. We realize that such kind of bottom-up level data may present different distribution if we choose alternate scale and this need to be further clarified. We also assume that each entry in our dataset is independently contributed, without taking the author of each contribution into account. Providing that some of the entries may be contributed by the same author, we may have a different result for our study.

For the future study, since we only look for the spatial autocorrelation at a national level this time by performing the global univariate Moran's I, a sub-regional dimension study can be conducted to compare the result to what we have right now. In addition, since the count number is a raw observation of the events, we may perform a further study by looking the specific rate of each province rather than just looking the raw count number. Another improvement would be that more social-economical factor and variable would be applied to robust the regression model. Last but not least, we could account for the author and content of each contribution to conduct a people-based and place-based study of this subject.

In retrospect, this chapter presents how VGI has been used to address a social inequity issue – street child begging and missing children in China. Using a representative sample of microblog posts as well as contextual information from a variety of sources, we are able to identify the spatial location of street children via the geo-tagged photos and text content from microblog. Topics from website provide location of missing and found children as well as when they are found and missed. With the intersection of VGI in Space as denoted in intersection "1" in Figure 29, we explore the spatial distribution of street children and missing children in China. Meanwhile, as denoted in intersection "2" in Figure 29, our study find out that VGI from these data source also contain valuable information which could help these missing children find their way home or help their parents rescue the street children. This includes the characteristics of children such as sex, age, and name. Also, missing and found information from VGI could be a valuable source to study missing children's mobility pattern as well.

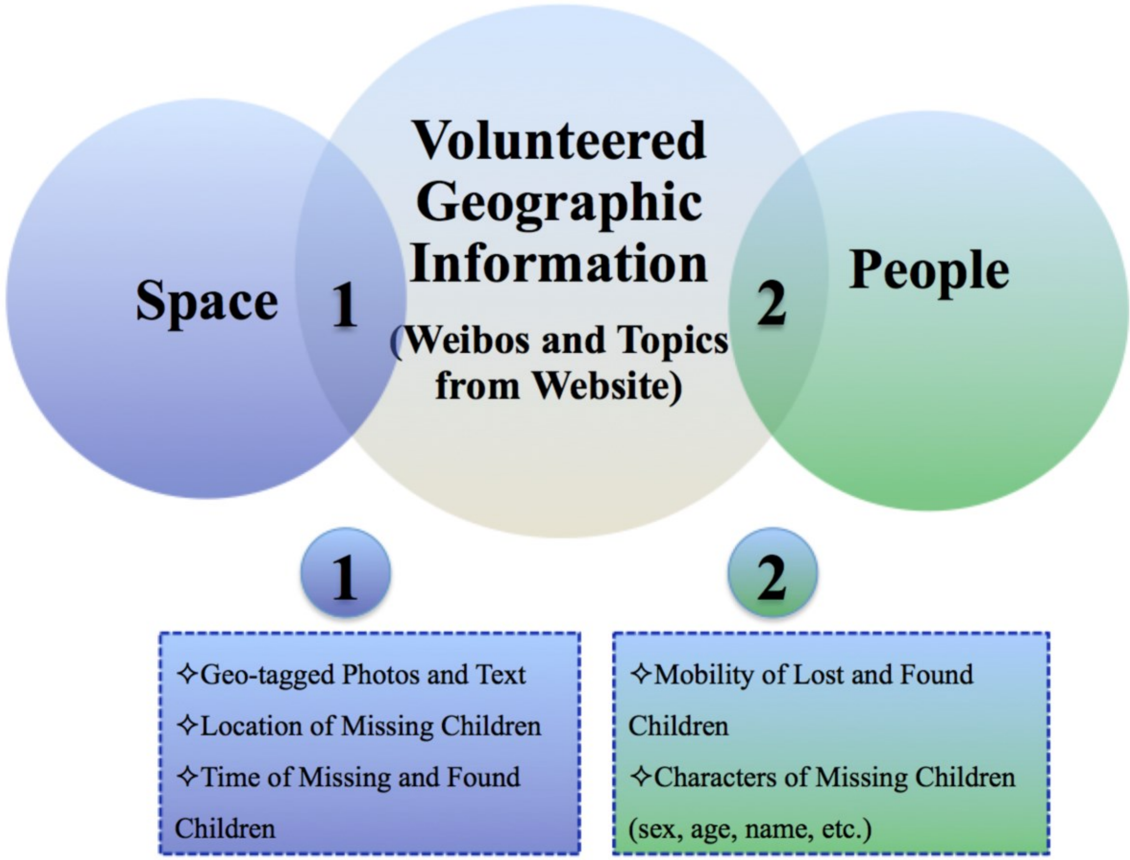


Figure 29 Framework to Study Street Child Begging and Missing Child

Chapter 6: Summary and Conclusion

6.1 Summary

Volunteered geographic information provide geographers a new gold mine to better understand our world. This dissertation first reviews the origin, type and current research of volunteered geographic information (VGI) in Chapter 2. In the same chapter I then examine space and place from a geographical perspective. We develop a conceptual framework linking VGI to *space, place and people* in chapter 2.3. In Chapter 3 we outline the mechanism on how to harvest geo-social media data that is widely used in this dissertation along with a subsequent methods of exploring the urban dynamic, leading to a case study of the pulse of a city within the Columbus, Ohio MSA using social media checkin data. This study offers insights of using VGI to study urban dynamics in space and place. We shift our lens on people and their exposed food accessibility to different place in Chapter 4 by exploring the applicability of geo-social media data on studying individual food environment. This chapter provides a novel method to use VGI as data source to study food-related issue on individual level and a 5-weekday period case study in Columbus, Ohio reveals some interesting finding on people's choice of food and their exposure to food environment. In Chapter 5 we present how VGI has been used to address a social inequity issue –street child begging and missing child, that has gained widespread attention on Chinese social media. Using a representative sample of microblog posts as well as contextual information from a variety of sources, this chapter examines the geographic distribution of street children in China as well as identifies how VGI are capable of empowering citizens to help those children find their home back.

6.2 Conclusion and future research

The context of geographic research has recently shifted from a data-scarce to a data-rich environment (Miller and Goodchild, 2014) and we believe the ubiquity of volunteered geographic information has the potential to fundamentally widen the angle for geographers to better understand our world. We understand that to accomplish this requires a series of development in new frameworks, new methods and new social practice. In this dissertation, we have focused on outlining three conceptual linkages with VGI and highlighted three corresponding case studies using geo-social media toward informing this emerging research area.

As this dissertation is written (early 2015), we realize that with the growing heterogeneity of VGI and its contributors, the quality, credibility, and uncertainty of VGI has also increased. This could be challenging to communicate with the public or for people who are used to working with authoritative data sources in a traditional top-down model (Haklay et al., 2014). Also, we believe the following two relevant trajectories are worth noting and need further attention.

First of all, we notice there is an ongoing fundamental change in the discipline of geography regarding the nature of data available to researchers, which leads to what some call *Big Data* (Miller and Goodchild, 2014). In fact, one interesting development during the past five years is that VGI is rapidly becoming a branch of the big data stream and is also merging into the big data deluge. Not surprisingly, as more and more big data with spatial and temporal stamps are available, boundaries between VGI and non-VGI are increasingly blurred day by day. Perhaps one day in the future we will use the term VGI and Big Data interchangeably, or will be largely indistinguishable from more conventionally-produced spatial data. Through our study, VGI has proved to be a valuable source of data for multiple exciting applications, but at the same time we have to keep in mind that VGI also raises profound legal, ethical, social, and political issues regarding its collection, retention, dissemination and use.

Second, the development of VGI and geospatial big data has important implications for GIS education. As part of the geospatial revolution, we believe VGI has become one indispensable patch of the new intellectual landscape for the emerging GIS themes, which have been categorized into six streams (Sui, 2015). These are also serving as a new guide for what should be taught in our GIS curriculum: a. geodesign, b. map stories, c. synthesis, d. critical GIS, e. Geogames and f. Platial GIS.

Another important point we have learned from our study is a paradigm shift from analysis to synthesis in GIScience. Former AAG president Thomas Baerwald (2010) pointed out the importance of synthesis in geography since geography as a interdisciplinary fields builds on major emphases in spatial analysis, human-environment interaction and place-based regional analysis that encourage communication and interaction. In fact, better integration and synthesis of diverse sources of geo-referenced information was a top priority during the early days of GIS development in the 1960s and 1970s. However, spatial analysis has been at the forefront in GIS during the past 40 years

(Griffith et al., 2013) with a primary focus on improving spatial analytic functions. The spatial analysis tradition often took a reductionist approach, focusing on individual layers to identify spatial patterns, rather than on the synthesis of multiple layers. We argue that our study of VGI and its linkage to *space, place and people* resonates well with the traditional spirit of geography in its quest to understand the multidimensional nature of the Earth's surface. Also, our attempt to integrate VGI with conventional geographic data is a synthesis at data level. Methodologically we apply both quantitative and qualitative methods to answer different questions and answer question in different ways. Indeed, the interdisciplinary nature of geography itself relies on a methodological synthesis in the mixing of qualitative and quantitative methods and we believe how to develop mixed-method in geography to cross the quantitative-qualitative chasm is a big challenge and we are optimistic of future development of this area.

References

Abebe, T. (2009). Multiple methods, complex dilemmas: negotiating socio-ethical spaces in participatory research with disadvantaged children. *Children's geographies*, 7(4), 451-465.

Agnew, J. (1987). *Place and politics: The geographical mediation of state and society.* Boston: Allen & Unwin.

Agnew, J. (2005). Space: place', In: Cloke, P., Johnston, R. (Eds.), *Spaces of Geographical Thought*. London, Sage, pp 81 - 96

Agnew, J. (2011). Space and place. In: Agnew, J., Livingston, D.N. (Eds.), *The Sage Handbook of Geographical Knowledge.* Los Angeles, CA: Sage, pp. 316-330.

Algert, J., Agrawal, A., & Lewis, S. (2006). Disparities in access to fresh produce in low-income neighborhoods in Los Angeles. *American Journal of Preventive Medicine*, 30(5), 365-370.

Álvarez, M., Pan, A., Raposo, J., Bellas, F., & Cacheda, F. (2008). Extracting lists of data records from semi-structured web pages. *Data & Knowledge Engineering*, 64(2), 491-509.

Andrejevic, M. (2009). *iSpy: Surveillance and power in the interactive era*. University of Kansas.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, 27(2), 93-115.

Anselin, L. (1998). *Exploratory spatial data analysis in a geocomputational environment.*

Geocomputation: A primer. Wiley, New York, 77-94.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. *Geographical Information Systems: Principles, techniques, management and applications*, (Eds.), P. Longley, M. Goodchild, D. Maguire, and D. Rhind. Cambridge: Geoinformation Int.

Apparicio, P., Cloutier, M. S., & Shearmur, R. (2007). The case of Montréal's missing food deserts: evaluation of accessibility to food supermarkets. International *Journal of Health Geographics*, 6(4).

Arribas-Bel, D. (2014). Accidental, open and everywhere: emerging data sources for the understanding of cities. *Applied Geography*, 49, 45-53.

Baerwald, T. J. (2010). Prospects for geography as an interdisciplinary discipline. *Annals of the Association of American Geographers,* 100(3), 493-501.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis. New York: Wiley.*

Baginski, J., Sui, D., & Malecki, E. J. (2014). Exploring the Intraurban Digital Divide Using Online Restaurant Reviews: A case study in Franklin county, Ohio. *The Professional Geographer,* 66(3), 443-455.

Ball, K., Crawford, D., & Mishra, G. (2006). Socio-economic inequalities in women's fruit and vegetable intakes: A multilevel study of individual, social and environmental mediators. *Public Health Nutrition*, 9(5), 623-630.

Ball, K., Timperio, A., & Crawford, D. (2009). Neighbourhood socioeconomic inequalities in food access and affordability. *Health & Place*, 15(2), 578-585.

Batty, M. (2010). The pulse of the city. *Environment and Planning B: Planning and Design,* 37(4), 575-577.

Batty, M., Hudson-Smith, A., Milton, R., & Crooks, A. (2010). Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS*, 16(1), 1-13.

Berg, N., & Murdoch, J. (2008). Access to grocery stores in Dallas. *International Journal of Behavioral and Healthcare Research,* 1(1), 22-37.

Berry, B. (1964). Approaches to regional analysis: A synthesis. *Annals of the Association of American Geographers*, 54, 2-11.

Block, D., & Kouba, J. (2006). A comparison of the availability and affordability of a market basket in two communities in the Chicago area. *Public Health Nutrition*, 9(7), 837-845.

Bodenhamer, D. J. (2013). *An exploration of deep maps.* http://thepoliscenter.iupui.edu/index.php/an-exploration-of-deep-maps. (last accessed 02.01.13).

Boyd, D. M., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of ComputerMediated Communication* 13, 210–230

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

Bromley, R. D., & Mackie, P. K. (2009). Child experiences as street traders in Peru: Contributing to a reappraisal for working children. *Children's Geographies*, 7(2), 141-158.

Bunge, W. (1971). *Fitzgerald: Geography of a revolution.* (Schenkman, Cambridge, MA).

Burdette, H., & Whitaker, R. (2004). Neighborhood playgrounds, fast food restaurants, and crime: Relationships to overweight in low-income preschool children. *Preventive Medicine Preventive Medicine*, 38(1), 57-63.

Casey, E. S. (1997). *The fate of place: A philosophical history*. Univ of California Press.

Chaix, B., Meline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., et al. (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.

Chen, X., & Kwan, M. (2012). Choice set formation with multiple flexible activities under space-time constraints. *International Journal of Geographical Information Science*, 26(5), 941-961.

Chen, X., & Clark, J. (2013). Interactive three-dimensional geovisualization of space-time access to food. *Applied Geography*, 43, 81-86.

Chen, X., & Yang, X. (2014). Does food environment influence food choices? A geographical analysis through "tweets." *Applied Geography*, 51, 82-89

Cheng, F., & Lam, D. (2010). How is street life? An examination of the subjective wellbeing of street children in China. *International Social Work*, 53(3), 353-365.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. *ICWSM*, 2011, 81-88.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *Plos One*, 5(11).

Chow, T. E. (2013). "We Know Who You Are and We Know Where You Live": A research agenda for web demographics. *In Crowdsourcing Geographic Knowledge* (pp. 265-285). Springer Netherlands.

Christensen, P., & James, A. (Eds.). (2008). *Research with children: Perspectives and practices*. Routledge.

Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area*, 42(4), 514-523.

Clarke, G., Eyre, H., & Guy, C. (2002). Deriving indicators of access to food retail provision in British cities: Studies of Cardiff, Leeds and Bradford. *Urban Studies*, 39(11), 2041-2060.

Craglia, M., Ostermann, F., & Spinsanti, L. (2012). Digital Earth from vision to practice: Making sense of citizen-generated content. *International Journal of Digital Earth,* 5(5), 398-416.

Cramer, H., Rost, M., & Holmquist, L. E. (2011). Performing a check-in: Emerging practices, norms and 'conflicts' in location sharing using foursquare. *In MobileHCI '11. ACM,* 57-66

Cresswell, T. (2004). *Place: A short introduction.* London: Blackwell.

Crutcher, M., & Zook, M. (2009). Placemarks and waterlines: Racialized cyberscapes in post-Katrina Google Earth. *Geoforum*, 40(4), 523-534.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data'and leveraging the potential of the geoweb. *Cartography and geographic information science*, 40(2), 130-139.

Curry, M. R. (2005). Toward a geography of a world without maps: Lessons from Ptolemy and postal codes. *Annals of the Association of American Geographers*, 95(3), 680-691.

De Venanzi, A. (2003). Street children and the excluded class. *International journal of comparative sociology,* 44(5), 472-494.

Doel, M. A. (1999). *Poststructuralist geographies: The diabolical art of spatial science*. Rowman & Littlefield.

Drewnowski, A., & Specter, S. (2004). Poverty and obesity: The role of energy density and energy costs. *The American Journal of Clinical Nutrition,* 79(1), 6-16.

Driskell, J., Meckna, B., & Scales, N. (2006). Differences exist in the eating habits of university men and women at fast food restaurants. *Nutrition Research*, 26(10), 524-530.

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

Elwood, S. (2008). Volunteered geographic information: Key questions, concepts and methods to guide emerging research and practice. *GeoJournal*,72(3), 133-135.

Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum*, 42(1), 6-15.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers*, *102*(3), 571-590.

Engler, N.J., Teresa Scassa, D. R. Fraser Taylor, (2014). Cybercartography and volunteered geographic information. *Modern Cartography Series*, Volume 5, (Chapter 4), pp. 43-57

Ehrenberg, R. (2012). The next epidemics will be tweeted: twitter posts tracked with official data during cholera outbreak. *Science News*, 181(4), 16-16

Fan, C. C., & Sun, M. (2008). Regional inequality in China, 1978-2006. *Eurasian geography and Economics,* 49(1), 1-18.

Feick, R. D., & Roche, S. (2010). Introduction (to special issue on VGI). *Geomatica*, 64(1), 5-6.

Ferriday, D., & Brunstrom, J. (2008). How does food-cue exposure lead to larger meal sizes? *The British Journal of Nutrition*, 100(6), 1325-1332.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Obersteiner, M. (2009). Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1(3), 345-354.

Gao, S., Li, L., Li, W., Janowicz, K., & Zhang, Y. (2014). Constructing gazetteers from volunteered big geo-data based on Hadoop. *Computers, Environment and Urban Systems.*

Ghosh, D., & R. Guha. (2013). What are we "tweeting" about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science* 40 (2):90-102.

Glanz, K., Sallis, S., Saelens, B., & Frank, L. (2007). Nutrition environment measures survey in stores (NEMS-S): Development and evaluation. *American Journal of Preventive Medicine,* 32(4), 282-289.

Glanz, K., Kristal, A. R., Sorensen, G., Palombo, R., Heimendinger, J., & Probart, C. (1993). Development and validation of measures of psychosocial factors influencing fat- and fiber-related dietary behavior. *Preventive Medicine*, 22(3), 373-387.

Goodchild, M. F., Egenhofer, M. J., Kemp, K. K., Mark, D. M., & Sheppard, E. (1999). Introduction to the Varenius project. *International Journal of Geographical Information Science*, 13(8), 731-745.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211-221.

Goodchild, M. F. (2009). The quality of geospatial context. *In Quality of context* (pp. 15-24). Springer Berlin Heidelberg.

Goodchild, M. F. (2011). Formalizing place in geographic information systems. In: Linda M. Burton, Stephen A. Matthews, ManChui Leung, Susan P. Kemp, David T. Takeuchi (Eds.), *Communities, Neighborhoods, and Health* (pp. 21-33). Springer New York.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), 231-241.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110-120.

Griffith, D. A., Chun, Y., O'Kelly, M. E., Berry, B. J., Haining, R. P., & Kwan, M. P. (2013). Geographical Analysis: Its first 40 years. *Geographical Analysis*, 45(1), 1-27.

Gupta, D., Spitzberg, B., Tsou, M., Gawron, M., An, L (2012). *Revolution in Social Science Methodology and Pitfalls.* Available on-line at http://mappingideas.sdsu.edu/publications/Revolution_final_version.pdf (last accessed 02.01.13).

Guy, C., & David, G. (2004). Measuring physical access to 'healthy foods' in areas of social deprivation: A case study in Cardiff. *International Journal of Consumer Studies,* 28(3), 222-234.

Hägerstraand, T. (1970). What about people in regional science? *Papers in regional science,* 24(1), 7-24.

Haining, R. (1993). *Spatial data analysis in the social and environmental sciences.* London & New York: *Cambridge* University Press.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B Planning and Design,* (37), 682-703.

Haklay, M., V. Antoniou, S. Basiouka, R. Soden, and P Mooney. (2014). *Crowdsourced Geographic Information Use in Government. Report to Global Facility for Disaster Reduction and Recovery (World Bank):* London, UK.

Hallett, L., & McDermott, D. (2011). Quantifying the extent and cost of food deserts in Lawrence, Kansas, USA. *Applied Geography*, 31(4), 1210-1215.

Hart, R. (1979). *Children's experience of place*. Oxford, Englind: Irvington.

Harvey, D. (1985). *The urbanization of capital: Studies in the history and theory of capitalist urbanization (Vol. 2)*. Baltimore: Johns Hopkins University Press.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Malden, MA: Blackwell.

Helling, A., & Sawicki, D. (2003). Race and residential accessibility to shopping and services. *Housing Policy Debate*, 14(1-2), 69-101.

Hill, L. L. (2000). Core elements of digital gazetteers: Placenames, categories, and footprints. *In Research and Advanced Technology for Digital Libraries* (pp. 280-290). Springer Berlin Heidelberg.

Holsten, J. (2008). Obesity and the community food environment: A systematic review. *Public Health Nutrition*, 12(3), 397e405.

Hunker, H. L. (2000). *Columbus, Ohio: A personal geography.* Ohio State University Press.

Jago, R., Baranowski, T., & Harris, M. (2006). Relationships between GIS environmental features and adolescent male physical activity: GIS coding differences. *Journal of Physical Activity & Health*, 3(2), 230-242.

Jeffery, R., Baxter, J., McGuire, M., & Linde, J. (2006). Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition & Physical Activity*, 3(2).

Jiang, B., & Miao Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer,* 67(2), 295-306

Kestens, Y., Lebel, A., Daniel, M., Thériault, M., & Pampalon, R. (2010). Using experienced activity spaces to measure foodscape exposure. *Health & Place*, 16(6), 1094-1103.

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*,331(6018), 719-721.

Knox, J. K., (2011). *Maneuvering global spaces by marketing local places: The process and practice of downtown revitalization in Columbus, Ohio* (Master Thesis, The Ohio State University).

Kwan, M. P., Casas, I., & Schmitz, B. (2004). Protection of geoprivacy and accuracy of spatial information: how effective are geographical masks? *Cartographica*, 39(2), 15-27.

Kwan, M. P., & Hong, X. D. (1998). Network-based constraints-oriented choice set formation using GIS. *Geographical Systems*, 5, 139-162.

Laraia, B., Siega-Riz, A., Kaufman, J., & Jones, S. (2004). Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Preventive Medicine*, 39(5), 869-875.

Larsen, K., & Gilliland, J. (2008). Mapping the evolution of 'food deserts' in a Canadian city: supermarket accessibility in London, Ontario, 1961-2005. *International Journal of Health Geographics*, 7(16).

Li, L., & Goodchild, M. F. (2012). Constructing places from spatial footprints. *In Proceedings of the 1st ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information* (pp. 15-21). ACM.

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science,* 40(2), 61-77.

Lin, W. (2013). When Web 2.0 Meets Public Participation GIS (PPGIS): VGI and spaces of participatory mapping in China. *In Crowdsourcing Geographic Knowledge* (pp. 83-103). Springer Netherlands.

Liu, S. B., & Palen, L. (2010). The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science,* 37(1), 69-90.

Lu, Y., & Liu, Y. (2012). Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. *Computers, Environment and Urban Systems*, 36(2), 105-108.

Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data. In Gold, M. (Eds.), *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press.

Martine, G., & Marshall, A. (2007). State of world population 2007: Unleashing the potential of urban growth. *In State of world population 2007: unleashing the potential of urban growth. UNFPA.*

Massey, D. (1991). A global sense of place. *Marxism today,* 35(6), 24-29.

McEntee, J., & Agyeman, J. (2010). Towards the development of a GIS method for identifying rural food deserts: Geographic access in Vermont, USA. *Applied Geography,* 30(1), 165-176.

McKinnon, R., Reedy, J., Morrissette, M., Lytle, L., & Yaroch, A. (2009). Measures of the food environment: A compilation of the literature, 1990e2007. *American Journal of Preventive Medicine*, 36(4), 124-133.

Miller, H. J. (2010). The data avalanche is here:Shouldn't we be digging? *Journal of Regional Science,* 50(1), 181-201.

Miller, H. (2007). Place-based versus people-based geographic information science. *Geography Compass*, 1(3), 503-535.

Miller, H. J. (2005). What about people in geographic information science. In Fisher, P., Unwin D. (Eds.), *Re-Presenting Geographic Information Systems*, John Wiley, 215-242.

Miller, H. J., & Goodchild, M. F. (2014). Data-driven geography. *GeoJournal*, 1-13.

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one*, 8(5), -64417.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q., & Vespignani, A. (2013). The Twitter of Babel: Mapping world languages through microblogging platforms. *PloS One*, 8(4).

Moore, L., & Diez Roux, A. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health*, 96(2), 325-331.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 17-23.

Morland, K., Diez Roux, A., & Wing, S. (2006). Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *American Journal of Preventive Medicine,* 30(4), 333-339.

Mulangu, F., & Clark, J. (2012). Identifying and measuring food deserts in rural Ohio. *Journal of Extension*, 50(3).

Openshaw, S., Wymer, C., & Charlton, M. (1986). A geographical information and mapping system for the BBC Domesday optical discs. *Transactions of the Institute of British Geographers,* 296-304.

Paez, A., Mercado, R., Farber, S., Morency, C., & Roorda, M. (2010). Relative accessibility deprivation indicators for urban settings: definitions and application to food deserts in Montreal. *Urban Studies*, 47(7), 1415-1438.

Panter-Brick, C. (2002). Street children, human rights, and public health: A critique and future directions. *Annual review of anthropology*, 147-171.

Pearce, J., Witten, K., & Bartie, P. (2006). Neighbourhoods and health: A GIS approach to measuring community resource accessibility. *Journal of Epidemiology & Community Health,* 60(5), 389-395.

Pickles, J. (Ed.). (1995). *Ground truth: The social implications of geographic information systems*. Guilford Press.

Powell, L., Slater, S., Mirtcheva, D., Bao, Y., & Chaloupka, F. (2007). Food store availability and neighborhood characteristics in the United States. *Preventive Medicine Preventive Medicine*, 44(3), 189-195.

Purves, R., Edwardes, A., & Wood, J. (2011). Describing place through user generated content. *First Monday*, 16(9). doi:10.5210/fm.v16i9.3710

Qualman, E. (2010). *Socialnomics: How social media transforms the way we live and do business.* New York, John Wiley & Sons.

Raja, S., Ma, C., & Yadav, P. (2008). Beyond food deserts: Measuring and mapping racial disparities in neighborhood food environments. *Journal of Planning Education and Research*, 27(4), 469-482.

Rana, S., & Joliveau, T. (2009). NeoGeography: An extension of mainstream geography for everyone made by everyone?. *Journal of Location Based Services*, 3(2), 75-81.

Reisig, V., & Hobbiss, A. (2000). Food deserts and how to tackle them: A study of one city's approach. *Health Education Journal*, 59(2), 137-149.

Relph, E. (1976). *Place and placelessness*. London: Routledge.

Rice, M. T., Paez, F. I., Mulhollen, A. P., Shore, B. M., & Caldwell, D. R. (2012). *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data.* Available on-line at: http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf (last accessed 01/06/2015).

Robillard, J., Johnson, T., Hennessey, C., Beattie, B., & Illes, J. (2013). Aging 2.0:

Health information about dementia on Twitter. *Plos One*, 8(7).

Sakaki, T., Okazaki, M., & Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering,* 25(4), 919-931.

Scassa, T. (2013). Legal issues with volunteered geographic information. *The Canadian Geographer/Le Géographe canadien*, 57(1), 1-10.

Scellato, S., Noulas A., Lambiotte R., & Mascolo C. (2011). Socio-spatial properties of online location-based social networks. *ICWSM' 11.* 329-336

Shaw, H. (2006). Food deserts: Towards the development of a classification. *Geografiska Annaler Series B: Human Geography*, 88(2), 231-247.

Shaw, S. L., Yu, H., & Bombom, L. S. (2008). A space‐time GIS approach to exploring large individual‐based spatiotemporal datasets. *Transactions in GIS*, 12(4), 425-441.

Simmons, D., McKenzie, A., Eaton, S., Cox, N., Khan, M., Shaw, J., et al. (2005). Choice and availability of takeaway and restaurant food is not related to the prevalence of adult obesity in rural communities in Australia. *International Journal of Obesity*, 29(6), 703-710.

Smith, C., & Morton, L. (2009). Rural food deserts: Low-income perspectives on food access in Minnesota and Iowa. *Journal of Nutrition Education and Behavior Journal*, 41(3), 176-187.

Smith, A., & Brenner, J. (2012). *Twitter use 2012. Pew Research Institute (technical report)* http://pewinternet.org/Reports/2012/Twitter-Use-2012.aspx Accessed 08.01.14.

Stamp, L. D. (1931). The land utilization survey of Britain. *The Geographical Journal,* 78(1), 40-47.

Strehl, T. (2011). The risks of becoming a street child: Working children on the streets of Lima and Cusco. *In editor?? Hazardous Child Labour in Latin America* (pp. 43-65). Springer Netherlands.

Sui, D. (2015). Emerging GIS themes and the six senses of the new mind: is GIS becoming a liberation technology?. *Annals of GIS*, 21(1), 1-13.

Sui, D. (2011). GIS: From space to place*? GeoWorld*, 24(12), 12.

Sui, D. Z. (2008). The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems*, *32*(1), 1-5.

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science,* 25(11), 1737-1748.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 234-240.

Tsou, M. H., & Yang, J. A. (2012). Spatial analysis of social media content (tweets) during the 2012 US republican presidential primaries. *In Proceeding of GIScience 2012.*

Tuan, Y. F. (1977). *Space and place: The perspective of experience*. U of Minnesota Press.

Tuan, Y. F. (1979). *Space and place: Humanistic perspective* (pp. 387-427). Springer Netherlands.

Turner, A. (2006). *Introduction to neogeography*. O'Reilly Media, Inc..

Twitter. (2013). http://www.statisticbrain.com/twitter-statistics/ (Accessed 08.31.14).

USDA (United States Department of Agriculture). (2009). Access to affordable and nutritious food: Measuring and understanding food deserts and their consequences. http://www.ers.usda.gov/Publications/AP/AP036 (Accessed 31.01.14).

Vossen, G., & Hagemann, S. (2010). *Unleashing Web 2.0: From concepts to creativity*. Elsevier.

Wainwright, J., & Barnes, T. J. (2009). Nature, economy, and the space-place distinction. *Environment and Planning D: Society and Space*, 27, 966-986.

Walker, R. E., Fryer, C., Butler, J., Keane, C., Kriska, A., & Burke, J. G. (2011). Factors influencing food buying practices in residents of a low-income food desert and a low-income food oasis. *Journal of Mixed Methods Research*, 5(3), 247-267.

Walker, R., Block, J., & Kawachi, I. (2012). Do residents of food deserts express different food buying preferences compared to residents of food oases? A mixed-methods analysis. *International Journal of Behavioral Nutrition and Physical Activity,* 9(41)

Wei, Y., & Ma, L. J. (1996). Changing patterns of spatial inequality in China, 1952-1990. *Third World Planning Review*, 18(2), 177-192

Widener, M. J., & Li, W. (2014). Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography*, 54, 189-197.

Wilson, M. W. (2012). Location-based services, conspicuous mobility, and the location-aware future. *Geoforum*, 43(6), 1266-1275.

World Bank. (2005). *World Development Report 2006: Equity and development*. Oxford University Press, Incorporated.

Wrigley, N., Warm, D., Margetts, B., & Whelan, A. (2002). Assessing the impact of improved retail access on diet in a 'food desert': A preliminary report. *Urban Studies,* 39(11), 2061-2082.

Yang, W., & Mu, L. (2014). GIS analysis of depression among Twitter users. *Applied Geography,* Available online

Yap, L. F., Bessho, M., Koshizuka, N., & Sakamura, K. (2012). User-generated content for location-based services: A review. In *Virtual Communities, Social Networks and Collaboration* (pp. 163-179). Springer New York.

Young, L., & Barrett, H. (2001). Adapting visual methods: action research with Kampala street children. *Area*, 33(2), 141-152.

Yue, Y., Lan, T., Yeh, A. G., & Li, Q. Q. (2014). Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, 1(2), 69-78.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7-33.

Zook, M., & Poorthuis, A. (2014). Offline Brews and Online Views: Exploring the geography of beer tweets. *In The Geography of Beer* (pp. 201-209). Springer Netherlands.