Geographies of Data:
Toward a Relational Socio-spatial Analysis of Geotagged Social Media Data

John Taylor Shelton

December 2015

A DISSERTATION

Submitted to the faculty of Clark University,
Worcester, Massachusetts,
in partial fulfillment of
the requirements for
the degree of Doctor of Philosophy
in the Graduate School of Geography

DISSERTATION COMMITTEE

James T. Murphy, PhD
Chief Instructor

Mark Davidson, PhD
Committee Member

Deborah Martin, PhD
Committee Member

Matthew Zook, PhD
Committee Member

ABSTRACT

This dissertation is an exploration of the potentials for utilizing geotagged social media data drawn from Twitter – one of many emerging sources of so-called 'big data' – for geographical research. The three papers that make up this dissertation examine how novel combinations of existing conceptual and methodological approaches from geography can be applied to datasets that are popularly understood as being revolutionary and requiring wholly new and previously unforeseen methods of analysis. Specifically, this dissertation seeks to combine the methodological approach of critical GIScience with the conceptual disposition of relational socio-spatial theory in order to explore a wider range of socio-spatial processes embedded in this data than is conventionally done in more popular or technically-oriented social media mapping projects. As such, this dissertation attempts to demonstrate how particular combinations of theory and methods can allow for more substantive insights to be drawn than is commonly thought possible due to the proliferation of relatively simplistic analyses of this data. This dissertation asks three key questions: (1) How have geographers and big data researchers conceptualized the spatiality of big data, and how do these conceptualizations constrain or enable the analysis of geotagged social media data? (2) How can existing frameworks for conceptualizing the multidimensionality of socio-spatial relations be usefully applied to the analysis of big data? (3) How can relational socio-spatial theory and critical GIScience be applied to big data in order to produce alternative ways of imagining urban spaces and places and the inequalities between them?

ACADEMIC HISTORY

Name: John Taylor Shelton **Date:** December 2015

Baccalaureate Degree:

University of Kentucky, Lexington, Kentucky
B.A., Geography and Political Science, December 2008
Minor in Appalachian Studies

Master's Degree:

University of Kentucky, Lexington, Kentucky
M.A., Geography, May 2011

Occupation and Academic Connection:

University of Kentucky, Kentucky Transportation Center
Research Associate, January 2015 – July 2015

University of Kentucky, Office of eLearning
Instructional Designer, July 2014 – December 2014

Clark University, Graduate School of Geography
Teaching and Research Assistant, August 2011 – December 2013

University of Kentucky, Department of Geography
Teaching and Research Assistant, January 2009 – May 2011

DEDICATION

For Emily, and for the Commonwealth.

ACKNOWLEDGMENTS

As the process of completing this PhD nears its end, it's something of a Herculean task to appropriately acknowledge all of those who have had a substantial role in getting me to this point. The number of people who have gone out of their way to make my life easier or more enjoyable over the past six-and-a-half years of graduate school is too large to properly enumerate, and some are surely left out here.

To begin, I would like to thank those who have contributed significantly to the development and completion of this dissertation. First, thanks go to my advisor, Jim Murphy, and committee members Mark Davidson and Deb Martin, for their various contributions to my development as a geographer over the past four years, and, ultimately, to this dissertation. An unending debt of gratitude goes to Matt Zook, for his more-often-than-not graceful shepherding of me through the highs and lows of graduate school. Without Matt's willingness to provide ample opportunities for intellectual growth and exploration, I can't imagine where I would be today. Ate Poorthuis deserves a similar level of appreciation, for his friendship and intellectual camaraderie have not only given me much to think about now and in the future, but have also directly enabled the work that makes up this dissertation. I cannot think of a more truly first-rate collaborator and team player than Ate. Mark Graham's contributions to Chapter 3 are also duly recognized – Mark has been a key collaborator and friend throughout the past six years. A great deal of thanks also goes to Matt Wilson, whose unbiased opinion and advice about navigating the world of academia I have often turned to in times of need, and just as often chosen to ignore. Though he won't receive any kind of 'official' credit for my completion of this PhD, his coaching and encouragement has been invaluable, and I have been happy to call him a close friend.

Next, it is worth noting all of those whose friendship has pushed me through this process, even at its most insufferable moments. At Clark and in Worcester, Renee Tapp, Alan Wiig, Kelly Kay, Bill Kutz, Alex Sphar and Padini Nirmal have been confidants and commiserators. The rest of my cohort and the entire CUGS community deserve some bit of recognition for putting up with my incessant talking about Kentucky throughout our time in Worcester, which was surely more than a little bit annoying. Garrett and Marianne FitzGerald deserve special thanks for their hospitality in Massachusetts, easing our transition by providing welcome reprieves from Worcester in the many fine eateries and pubs of the greater Boston area and on the beaches of the North Shore and Cape Cod. In Kentucky (and, now, places beyond), Patrick Bigger, Jairus and Ali Rossi (and Nica and Frank), Brian Grabbatin, Jon Otto, Nate Millington, Mitch Snider, Hugh Deaner, Chris Van Dyke, Tim Brock, Candice Wallace, Kenny Stancil, Lauren Martin, Oliver Belcher, Stephanie Simon and Michael Marchman and a number of others have been some of the best friends and role models I could ask for – you all make me proud to be a Kentucky geographer. I would be remiss, however, in not specifically mentioning Oliver and Marchman for their role in bringing me to geography in the first place. It was their April 2006 "Kentucky Educator's Pledge" that let me know geography was where I wanted to be…the rest, they say, is history.

Further thanks go to my parents, sisters, brothers-in-law, nephew, niece and extended family for having helped prepare me for this journey, even if they could never fully grasp what it was like to undertake a PhD. And, as is customary with these types of things, the final but most important thanks goes to my partner, Emily, for her wherewithal through the past nine-and-a-half years (but especially the last four) has helped push me to this point. Without her presence – and many, many long walks with Mali – I know that I wouldn't be here now, nor would I want to be.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction

I. Introduction to the Dissertation

This dissertation is an exploratory study of the potentials for utilizing emerging sources of so-called 'big data' – in particular, geotagged social media data – for geographical research. The three papers that make up this dissertation look at how existing conceptual and methodological frameworks from critical geography can be applied to datasets that are popularly understood as being revolutionary in that they require wholly new and unforeseen methods of analysis, free of the confines of theorizing (cf. Anderson 2008). More specifically, this dissertation builds upon the longer history of critical GIScience scholarship within geography, especially the tradition of critical GIScience which seeks to not only critique these socio-technical practices through a deconstruction of their effects in the world, but also critique these ways of producing knowledge through a more practical engagement, so as to "open up other ways of knowing" (Crampton 2010: 15) *through* these technologies and techniques. In applying the sensibilities of critical GIScience to the study of geotagged social media data, the work in this dissertation also builds from Crampton et al's (2013) call to go 'beyond the geotag' in studies of geotagged social media data in order to consider a range of more complex social and spatial processes than are conventionally represented by focusing on only the geographic coordinates attached to such data. In so doing, this work also draws on work from what might broadly be understood as relational socio-spatial theory, so as to demonstrate how novel combinations of these heretofore unrelated conceptual and methodological approaches can not only be operationalized with respect to these new sources of data, but also yield alternative insights from such data than are often achieved through more conventional technical approaches.

II. Situating Big Data in Geography

'Big data' remains an ill-defined and nebulous term throughout society, with definitions shifting in such a way as to behoove whomever it is defining the term in a given moment. Definitions range from Laney's (2001) somewhat older '3 Vs' – volume, variety and velocity – definition to Kitchin's expansion of these three traits to also include characteristics of exhaustiveness, fine-grained in resolution, indexicality, relationality and flexibility (Kitchin 2013). Other definitions focus exclusively on the 'bigness' of the data, such as the notion that big data is any "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyinka et al 2011), or to use the oft-cited, more colloquial version, 'any data that cannot fit into an Excel spreadsheet'. But, as Graham and Shelton (2013) have argued, "Whatever exactly big data is, it appears as though something important has changed" (256). The expectation that the global market in big data services would reach $16 billion annually in 2014 is a testament to this apparent shift and the valorization of big data, more generally (Press 2013).

Much of the hype around big data rests on the "widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy" (boyd and Crawford 2012: 663). But as Graham and Shelton argue further, "in the rush to start using big data, many have neglected questions about where or how it might be integrated into preexisting structures of scholarly knowledge production" (Graham and Shelton 2013: 256). This has been especially true of a range of scholars from outside the social sciences who have sought to leverage these new sources of social data to uncover what they see as fundamental laws of human society, bringing the approaches and disposition of the natural sciences to the study of social life (cf. Pentland 2014). Indeed, a significant portion of this work on 'social physics' addresses explicitly geographic questions around issues of mobility, environmental perception, and urbanization, among other things (Bettencourt and West 2010; Noulas et al 2012; Salesses et al 2013; Batty 2013).

It is important to note, however, that for all of the talk about big data as allowing for a revolution in our ways of understanding the world around us, many of the things that qualify as 'big data' aren't accessible to the typical researcher, much less to the average citizen or casual consumer. Big data includes everything from the log of clicks one makes within an internet browser, the history of purchases on one's credit card and personalized medical records to various digital traces of one's presence through the city thanks to mobile phone tracking and sensor networks embedded in the built environment. In part because of the very 'closed' nature of much of what counts as big data, geographers and other social researchers have been drawn to data produced through social media platforms such as Twitter, Flickr, Instagram and Foursquare, due to their relative openness and accessibility, often discussed under the umbrella of ‘volunteered geographic information’, ‘neogeography’ or the ‘geoweb’ (Goodchild 2007, 2009; Graham 2010; Elwood et al 2012). So while data from these sources are used in a significant proportion of research within geography that mobilizes the framing of big data, it is important to keep in mind that this data represents only a relatively small subset of that which might be understood as big data. For this reason, this dissertation seeks to simultaneously engage with broader discourses around ‘big data’, while grounding a critique of big data and the dominant ways of thinking about it through the utilization of one specific type or source of big data – geotagged social media data drawn from Twitter.

Indeed, while this dissertation explicitly seeks to build on the longer history of work within critical GIScience in order to shed light on the promises and pitfalls of emerging big data sources, it is in relation to questions of data that this work diverges from prominent traditions of critical GIS in a major way. Many of the seminal papers demonstrating the potential of critical and participatory approaches to GIS were based on very deliberate processes of data collection and analysis in line with a commitment to more democratic, and often explicitly feminist, ways of producing knowledge and solving social problems (cf. Kwan 2002; Pavlovskaya 2002; Elwood 2006; Knigge and Cope 2006). In contrast, this dissertation’s starting point is an attempt to make sense out of data that wasn’t originally intended for social research at all, much less any particular kinds of interventions into local problems from a broadly ‘critical’ perspective. That being said, data drawn from social media platforms and other sources of user-generated information has already been put to use for a variety of purposes within geographic research. From studies of food accessibility (Ghosh and Guha 2013; Chen and Yang 2014; Widener and Li 2014; Zhai et al 2015) to cultural geographies of food (Zook and Poorthuis 2014; Poorthuis et al, forthcoming), religion (Zook and Graham 2010; Shelton et al 2012; Wall and Kidnark 2012) and language (Watkins 2012; Graham and Zook 2013; Graham et al 2014a), as well as the geographies of social movements (Hemsley and Eckert 2014), cognitive/perceptual geographies and socio-spatial imaginaries (Kelley 2013; Xu et al 2013; Han et al 2015), processes of socio-spatial diffusion (Tsou et al 2013), understanding inequality and uneven development as it relates to digital information (Crutcher and Zook 2009; Graham and Zook 2011; Graham et al 2014b), as well as place identities (Cranshaw et al 2012; Stefanidis et al 2013; Feick and Robertson, forthcoming), urban planning (Ciuccarelli et al 2014) and even more technical GIScience work on overcoming longstanding issues related to the granularity of social data and attempts at small area estimation (Lin and Cromley 2015; Longley et al 2015).

Despite the range of applications of this data to an assortment of existing and emerging subfields within geography, so too have emerged a series of critiques of this data and the larger discourse around big data in the social sciences. For example, Bowker (2014) and boyd and Crawford (2012), among others, have argued that many of those celebrating these new sources of data have pushed aside questions about the fundamentally theory and value-laden nature of big data research, in spite of pervasive claims to the contrary. Similarly, Barnes (2013) and Barnes and Wilson (2014) have argued that far from being a new, revolutionary approach to social research, many of these quantitative and computational approaches have a much longer, if oft-forgotten, history within geography and the social sciences. For this reason, Miller and Goodchild (2015) argue that while "data-driven geography may seem revolutionary, in fact it may be better described as evolutionary since its challenges have long been themes in the history of geographic thought and the development of geographical techniques" (450). Meanwhile, others have pointed towards the persistent relevance of more conventional 'small data' approaches and qualitative research methods in the era of big data (Kitchin and Lauriault 2015).

A further subset of these critiques, however, come directly out of traditions of critical GIScience and critical quantitative geography, in effect rearticulating the early- and mid-1990s critiques of GIS with respect to big data. These early critiques of GIS, as noted by Schuurmann (2000), kept a kind of distance between themselves and the technology at the center of critique. From Smith's (1992) recognition of the role of new geospatial technologies in enabling the Gulf War, to warnings of the erosion of privacy (Goss 1995; Curry 1997), to Pickles' (1995b) arguments about the fundamentally anti-democratic nature of GIS and Taylor's (1990) assertion that GIS was more concerned with 'geographical facts' than 'geographical knowledge', leading to "a return of the very worst sort of positivism, a most naïve empiricism" (212), there was a pervasive sense that GIS was a threat to the more radical, critical or even humanistic traditions within geography. This sensibility has been carried forward in recent scholarship within critical GIS that engages with big data, for instance, focusing on how notions of privacy are changing with the increasing production of user-generated geographic information online (Elwood and Leszczynski 2011; Leszczynski, forthcoming) or critiquing claims to democratization (Leszczynski 2012; Haklay 2013). Similarly echoing earlier critiques of GIS, Thatcher (2014) sees a fundamental danger to utilizing these new sources of data due to their imbrications within systems of capitalist political-economic exploitation:

> "when academic researchers accept the resulting data as meaningful, they are fundamentally accepting an epistemological framework of knowledge structured through capitalist imperatives. When social science researchers use big data produced and maintained by private corporations, they tacitly accept the incentives and requirements that shaped the information…they are accepting not only that the object of study can be found within the 140-character limitation of the tweet but that their ability to access the information and the form it takes will be controlled by the privately owned Twitter Incorporated" (1772-1773)

From a somewhat similar position, Wilson (2015) argues against seeing, and particularly *using*, these new sources of data *as such*. In a recent commentary, he writes, "I specifically intend to push back on the proliferation of studies that propose to utilize social media as 'big data' evidence. I argue that we should lean more on the notion that social media are phenomena and less on the notion that social media are evidence of phenomena. In other words, the aggregation of social media as big data is not necessarily social science data" (Wilson 2015: 346).

While many of these criticisms are difficult to dispute, given the fact that big data has *already* been reconfiguring the theories and practices of social research across a range of scholarly and applied fields for close to a decade (cf. Savage and Burrows 2007), it seems more useful to attempt to understand how big data can be more usefully integrated with existing concepts and methodologies in human geography, rather than simply dismissing these developments for their obvious shortcomings. Indeed, human geography has much to contribute to the discourse around, and use of, big data for social research, just as big data offers new potentials for human geography. And while geographers have made important contributions to critiques of big data (see especially the forum in *Dialogues in Human Geography* 2013, as well as Kitchin 2014a), relatively little work within the discipline has thusfar attempted to bridge the divide between these emergent critiques of the increasingly-mainstream approach to big data and the actual analysis of such data sources according to a more geographically-oriented, context-sensitive approach. As such, this dissertation seeks to bridge the divide between these two camps, in much the same way as GIS was gradually intermingled with critical human geography in the later-1990s and early-2000s in the form of critical GIScience, as opposed to the more oppositional stance between the two somewhat earlier on.

III. Conceptual Framing

Human geographers have thus far seemed to be largely reticent to employ the analysis of big data in their research, at least partially due to the aforementioned symmetries between the emerging approaches of a transdiciplinary approach to 'data science' and earlier developments around GIS. But even when big data, especially in the form of geotagged social media data, has been analyzed by geographers or others with a particularly geographical bent, much of this work has tended to utilize a fairly simplistic understanding of space and socio-spatial relations that neglects larger developments within socio-spatial theory over the last few decades, an argument encapsulated in Crampton et al's (2013) call for such analyses to go 'beyond the geotag' (see Chapter 2). In order to address these shortcomings, this dissertation seeks to produce a novel combination of existing, but otherwise disparate, conceptual and methodological approaches so as "to think of new epistemologies that do not dismiss or reject Big Data analytics, but rather employ the methodological approach of data-driven science within a different epistemological framing that enables social scientists to draw valuable insights from Big Data that are situated and reflexive" (Kitchin 2014b: 9-10). Building on the epistemological and methodological stance of critical GIScience and integrating a concern with a more flexible spatial ontology drawn from relational socio-spatial theory, this dissertation seeks to bridge the gap between critical geographers hesitant to engage with 'the new quantitative revolution' (Wyly 2014) offered by big data, and those data scientists whose technical acumen is divorced from a geographically-situated understanding of the social processes their analytical tools seek to unravel.

A. Critical GIScience

In many ways, contemporary human geographers' reticence to integrate big data into their existing research programs mirrors earlier divisions within the discipline between social theorists and quantitative methodologists or GIScientists (cf. Pickles 1995a;

Schuurman 2000; Barnes 2013; Barnes and Wilson 2014). While earlier calls for human geographers to untangle the historically contingent associations between these methods, positivist epistemologies and reactionary politics and engage more directly with the use of these more technically-oriented approaches are clearly relevant to the question of big data (Sheppard 2001; Wyly 2009, 2011), the use of big data sources by more critical human geographers has been minimal. Whether this failure to engage across boundaries is the result of a persistent assumption that these data sources are inextricable from the hyper-positivist discourses that have thus far surrounded them, or simply because of the technical difficulty in accessing this data, it is fair to say that human geographers have not taken full advantage of the range of new data sources available to them and explored the range of alternative ends to which this kind of data can be used.

Given the similarities between contemporary debates around big data and earlier debates within geography around the growth of GIS, it is fitting to draw on an ever-growing literature around critical GIScience when discussing the potentials of alternative approaches to big data. An outgrowth of the earlier 'GIS and Society' debates of the early and mid-1990s, which saw a largely acrimonious relationship between GIScientists and social theorists, critical GIScience seeks a synthetic approach that brings together the technical capabilities of GIS and quantitative methods with a sensitivity to the broader social context in which these tools are situated. In other words, critical GISscience represents an attempt to overcome what was seen as a persistent problem of insiders and outsiders within critiques of GIS. And while a significant undercurrent of externalist critiques of GIS and associated technologies remains, this dissertation draws more specifically on traditions of critical GIScience scholarship, such as qualitative GIS, that "preserves much of the language and practices of conventional GIS methodology" while being "attuned to social theory" in order "to address the earlier critiques of the technology" (Preston and Wilson 2014: 513; see also Knigge and Cope 2006; Cope and Elwood 2009).

Indeed, the range of work that falls under qualitative GIS has demonstrated that rather than being fundamentally quantitative and positivist in origin and nature, GIS is a flexible technology that has long held the potential to integrate and represent more grounded, qualitative data in tandem with non-positivist epistemologies (Kwan 2002; Pavlovskaya 2006), which has become all the more true with the rise of new forms of spatial data production and representation that are more widely accessible than earlier iterations of GIS (Goodchild 2009; Warf and Sui 2010; Elwood and Leszczynski 2013). The use of new geospatial technologies and location-based services means that the spatial data being put to use in critical GIScience scholarship is substantially wider in scope and scale than was previously the case. Given that the geotagged social media data being analyzed in many cases comprises hundreds or thousands of unique users of varying backgrounds, this data offers a potentially wider range of voices and perspectives than would be possible through more conventional means. That being said, the expansiveness of the data presents challenges both in terms of its size, as well as its fundamentally unstructured nature, making it much more difficult to discern context and meaning from text. And while the later portions of this dissertation demonstrate some of the utility of such qualitative analysis as applied to targeted subsets of geotagged tweets, this work also points towards the need for much more thorough and sustained engagements with social media as source data for qualitative analysis.

In short, the existing body of work within critical GIScience demonstrates the potential of engaging directly with the analysis of otherwise maligned big data sources without sacrificing a critique of this data and the discourses that have been built around it. As Elwood writes, "the conceptual and practical interventions of critical GIS…do not render GIS unproblematic, but rather, they open up these limits for critique and disruption through purposeful and reflective engagements with GIS" (2010: 55). But despite the significant progress in demonstrating how GIS can be utilized in tandem with critical social theory and more qualitative, context-sensitive methodologies, one area which has remained relatively unexplored within this work is the conceptualization of space and socio-spatial relations promoted by conventional GIS approaches, and how such conceptualizations might be changed to allow for a more robust understanding of spatiality. Even though much of the earlier critical GIScience literature remarked on the need for such a rethinking of space as represented through GIS (cf. Rundstrom 1995), relatively little headway has been made in this respect (Goodchild 2006). Even when more critical GIS work has departed from the GIS orthodoxy in order to argue for a more robust attention to questions of temporality and mobility, this work tends to couch such a shift as a de-emphasizing of space, rather than as a reconceptualization of space itself (cf. Kwan 2013). In order to address this persistent challenge within critical GIS, as well as the more contemporary issue around the limited insights drawn from the mapping of social media data, this dissertation seeks to mobilize these methods in concert with the conceptual approach of relational socio-spatial theory in order to demonstrate alternative modes of engaging with big data derived from social media platforms.

B. Relational Socio-spatial Theory

While some of the most problematic mappings of social media data come from those outside the discipline of geography, even when this data has been put to use by geographers there has been a tendency to utilize a relatively simplistic understanding of space and socio-spatial relations that is not in tune with broader trends in socio-spatial theory over the last two decades. Because these individual data points come with only a single pair of latitude and longitude coordinates attached to them, it is often assumed that this is the full extent to which this data is infused with geography. But, as Crampton et al (2013) highlight, this 'spatial ontology of the geotag' has elided the complexity of socio-spatial relations as they are embedded in this kind of social data, especially in more relational forms. As more conventional GIScience approaches have dominated this kind of research, a framework that privileges Cartesian, geometric understandings of space has stripped this data of the rich historical and geographical context in which it is embedded (cf. Sheppard 1995; Barnes 2013; Wyly 2014). It's important to note that while there's nothing inherently 'wrong' with this Cartesian spatial ontology – indeed, some element of Cartesian thinking is necessary within contemporary geographic information systems – this perspective is incredibly partial, limiting both the kinds of questions that can be asked of this data and the kinds of answers that might be found.

In order to overcome the limitations of the spatial ontology of the geotag and the associated Cartesianism of GIS, this dissertation seeks to demonstrate the utility of thinking space *relationally*, so as to highlight a more diverse range of social and spatial processes that can be analyzed using geotagged social media data. Drawing especially on Doreen Massey's (1991) early formulations around a 'global sense of place', relational socio-spatial theory conceives of space as networked, fragmented and processural, rather than as a kind of fixed container with defined boundaries and characteristics, such as single points or the more-or-less arbitrary Census-defined areal units typically used for these kinds of analyses. From reconceptualizations of globalization (Amin 2002) to a new focus on mobility as a fundamental, defining characteristic of contemporary life (Sheller and Urry 2006), a key tenet of this approach has been an inversion of Tobler's so-called 'first law of geography' – that all things are related, but near things are more related than far things. Instead, relational approaches suggest that "we cannot assume that local happenings or geographies are ontologically separable from those 'out there'" (Amin 2002: 386). By focusing on the social relations that recursively produce space and are in turn influenced by it, rather than simply privileging proximity in absolute, Cartesian space, Amin argues that we can begin to see "a subtle folding together of the distant and the proximate" (2007: 103). As social processes are more and more spatially extensive, owing at least in part to the increasing prevalence of information and communication technologies, our spatial categories similarly need to evolve so as not to assume universal connections between social activities or processes and the locations on the earth's surface at which they occur.

While much of this 'relational turn' in socio-spatial theory has been seen as running counter to work within a Marxian tradition of political-economic geography, especially as it relates to the concept of scale (cf. ongoing debates in Amin and Thrift 2002; Allen and Cochrane 2007; MacLeod and Jones 2007; McFarlane 2011; and Brenner et al 2011), more recent engagements with Henri Lefebvre and his understanding of 'planetary urbanization' have provoked a similar tendency to rethink accepted categories of socio-spatial analysis in favor of a more relational approach (Brenner 2013; Brenner and Schmid 2014). Similarly, work by Graham and Marvin (2001), Jessop et al (2008), McCann and Ward (2010) and Pierce et al (2011) has demonstrated the potential of a simultaneously relational *and* territorial understanding of socio-spatial relations as applied to questions of urban politics, producing much more nuanced understandings of everything from the role of urban infrastructure in exacerbating uneven geographical development, how urban policies get produced in particular places and circulated through networks, or how different regimes of governance are produced and enacted.

C. Combining Critical GIScience and Relational Socio-spatial Theory

By explicitly drawing on relational socio-spatial theory, this dissertation attempts to address a longstanding gap within critical GIScience research writ-large, and especially more recent discussions around new forms of 'volunteered geographic information'. As mentioned above, the earliest formulations of a GIS and Society research agenda – which can be seen at least as a direct precursor to, if not actually a part of, critical GIScience – prominently featured discussions around the necessity of including non-Cartesian or non-Euclidean spatialities within GIS (cf. Rundstrom 1995; Sheppard 1995; Yapa 1998). But as a critical GIScience matured through the integration of these critiques into the actual 'doing' of GIS with a critical orientation, this focus on spatiality was largely lost, perhaps due in part to the difficulty of 'solving' this problem. That is, integrating relational spatialities into GIS is something of a persistent meta-problem; GIS and relational understandings of space can be held in productive tension with one another, but the necessity of locating each object within a Cartesian coordinate system in order to represent phenomena in a computerized map remains a substantial barrier to producing any kind of 'final' solution to this matter.

So instead of experimenting with different ways of integrating this concern with relational space into the practice of critical GIS, much of the seminal work within critical GIScience focused more directly on the social context around, and political implications of, the use of GIS. This work particularly emphasized facilitating the use of GIS by otherwise marginalized populations so as to translate their knowledge into more-or-less 'legible' representations, and interrogating the tensions emerging from such translations (Ghose 2001; Elwood 2002; Kwan 2002; Elwood and Leitner 2003). Though this work excelled at demonstrating the potential of GIS to promote alternative understandings of social problems on behalf of marginalized communities, including the articulation of alternative claims on neighborhood spaces, this work didn't really question the received spatial categories at work in the construction of these neighborhoods by integrating an alternative understanding of space into such an analysis.

This elision of spatiality as a central concern for critical GIScience has similarly been the case in more recent discussions around the newly-enabled shift of geographic information production from experts to amateurs, commonly known under the label of 'volunteered geographic information' (or VGI). While the valorization of VGI by GIScience luminary Michael Goodchild (2007) itself harks back to the early GIS and Society agenda and its emphasis on integrating 'unauthoritative' sources of data into GIS research (cf. Miller 1995), this more recent work has still left behind an explicit attention to spatiality. For instance, Elwood's (2008) articulation of critical GIS-inspired research agenda for VGI focuses largely on the inputs of data production and the means to which this data can be put, without attention to the elements of the data or the methods of analysis that might shape these outputs. Similarly, while Dodge and Kitchin (2013) take a broadly relational ontology in interrogating VGI, they are focused more on the ontology of discrete objects – i.e., what things should or shouldn't be included in crowdsourced mapping projects – rather than the spatial ontologies implicit in the production of this data. Even when relational space has found its way into discussions of VGI, such as in Warf and Sui's (2010) arguments about the possibilities of alternative epistemologies and ontologies enabled by these new forms of data, the potentials are interpreted as being more about the *production* of geographic information outside the bounds of conventional desktop GIS' tabular data structures in accordance with alternative epistemologies and ontologies, rather than in the *analysis* of this data through a relational lens. That is, relational space remains a kind of proposition or challenge for VGI, rather than something that is actively engaged with and demonstrated through the analysis of this data.

As it relates to the study of geotagged social media data, combining such a relational (or simultaneously relational-territorial) approach with critical GIScience is useful because it doesn't overlook the range of qualitative and quantitative approaches that might allow one to better understand the context and meaning such big data, whereas those approaches which perpetuate a simplified spatial ontology tend to simply plot data points in Cartesian space in order to identify those places with most dense concentrations. Relational socio-spatial theory allows us to recognize that people's social lives and spatial practices are expressed in ways that can't remotely be captured by either treating individual data points as isolated, or by dealing in a priori aggregate areal units that may have very little correspondence whatsoever to the underlying social processes. This provides a more complex mapping of socio-spatial relations by drawing attention to a wider range of possible explanations for spatial patterns, pointing towards the fundamental interconnection of spaces and places that might typically be thought of as internally coherent and self-evidently defined.

For instance, a relational understanding of space helps to formulate methods of analysis that don't simply look for the mere presence or absence of data points in a particular place, but dig deeper to show how the extremely dense data shadows in some locations are intimately connected to the scarcity of data in other places. As demonstrated most clearly in Chapter 4, this can be due to the effects of public policy decisions that necessitate residents of poorer neighborhoods traveling to other places within the city for school, work and recreational activities, where the production of user-generated content tends to be more prevalent. In other words, a more simplistic analysis might simply note the lack of data in poorer neighborhoods and conclude that this data fundamentally underrepresents the populations that live in those neighborhoods, without recognizing that these individuals are actually substantive contributors to the overall density of data in the city, just not in their 'home' neighborhoods. So even if relational or multidimensional understandings of space can never be fully integrated into the default ways we engage with GIS, allowing relational understandings of space to guide our analyses and interpretations offers significant potential for reconsidering the value of these new forms of data, as well as for developing alternative understandings of urban socio-spatial dynamics based on empirical observations.

IV. Data Collection and Methods

Although this dissertation as a whole focuses on questions of theory and methodology, some initial discussion of method is appropriate at this time. In particular, it is important to outline some of the particulars of the data that serves as the centerpiece for this project as a whole, addressing questions of how it is collected, what kinds of metadata are associated with each individual data point, and how these pieces of metadata are the cornerstone of building an analysis of this data that isn't limited solely to each individual point's geographic coordinates, as well as some of the general analytical approaches adopted throughout the three papers.

In general, social media data from platforms like Twitter can be understood as a kind of 'data exhaust', the somewhat unintentional byproduct of intentional actions on the part of users of the given platform. So while an individual must choose to tweet, this choice comes with a number of secondary ramifications, one of which being the inclusion of that user's behavior into larger datasets of social media activity that can be archived and accessed at a later time. While this research is concerned with the explicitly geographic dimensions of tweeting activity, and thus focuses only on tweets themselves, the act of following and being followed by other users, as well as 'retweeting' or 'favoriting' another user's tweets are similarly catalogued and can be used to understand processes of social networking and how ideas diffuse through these networks online (Miller 2011). The use of this data, however, is limited both by the difficulty in understanding the data out of the context of the particular social practices that take place within this and other similar platforms, as well as biases in the data (Crawford 2013). So, while some see this digital data as enabling the *prediction* of offline social processes or events, such as presidential elections or the stock market, because of these biases it is perhaps more realistic to conceive of this data as a kind of real-time archive of social – and, in many cases, spatial – behavior, at what Miller (2011) calls "a huge scale", albeit with many persistent gaps and errors. With roughly 500 million tweets created each day, or around 200 billion tweets per year, by an estimated 316 million monthly active users[1], even analyzing only those tweets with explicit geotags – estimated to be anywhere less than 5% of all tweeting activity – leaves researchers with millions of artifacts of social behavior that can be pieced together relatively quickly and easily, compared with more conventional methods one might use to gather so many discrete pieces of information.

The data for this dissertation was collected through DOLLY[2], a project of the University of Kentucky's New Mappings Collaboratory, developed by Ate Poorthuis and Matthew Zook. DOLLY works to collect data from Twitter's Streaming API, cataloguing this data on-the-fly so as to allow for retrieval of data at a later point in time. While many smaller-scale studies of Twitter data have only been able to collect data on a limited number of topics during a defined period of time, DOLLY instead collects each and every geotagged tweet created worldwide and stores them in a searchable database, allowing for a greater level of flexibility in identifying potential

---

[1] All data approximate as of June 30, 2015, available from https://about.twitter.com/company.
[2] Data On Local Life and You. For more information, see http://www.floatingsheep.org/p/dolly.html.

topics of interest and queries that sufficiently capture a given issue as represented on Twitter. In the three years since its inception in July 2012, DOLLY has indexed over 11 billion geotagged tweets worldwide.

Figure 1.1: The Anatomy of a Tweet

Source: Poorthuis et al (forthcoming)

Key to DOLLY's architecture is that DOLLY not only collects the locations of tweets, but also a range of other metadata attached to the tweet that is rarely visible to the average Twitter user. While Figure 1.1 shows how some of these pieces of information are represented within the Twitter web interface – such as the name of the user, their Twitter handle or username, the time and date the tweet was created, as well as the text of the tweet and any associated hashtags, links or attachments – there are also a range of other pieces of information that aren't so easily accessible. For example, DOLLY also collects a unique numerical user identification number that allows for the collection of data about individual users, even if they happen to change the other identifying information about themselves, such as their username, as well as the 160-character biography attached to each profile, the total number of tweets that user had created at the time of a given tweet, the number of other accounts that they follow and are followed by. Each of these different fields, shown in Figure 2 as they would be seen after being downloaded and opened in Microsoft Excel, are searchable within the DOLLY interface using boolean logic. DOLLY even takes the location of the geotag and parses the latitude and longitude coordinates into different areal classifications, from the country to the state or administrative region, and in the United States, all the way down to the county and even census tract level. This automatic process of cataloguing each tweet allows for a more targeted approach to data extraction by specific locations.

But even the most fundamental piece of metadata – the geotag – isn't entirely straightforward. Twitter allows for different kinds of geotagging based on the user's preferences. This allows for some fuzziness to be introduced, should the user wish to share only a more general location, like the name of a city – 'Lexington, Kentucky' – or a particular neighborhood or landmark, or even, in some cases, tag their tweets to a location at which they are not currently present[3]. Each of these 'places' carries differing levels of specificity in the latitude and longitude coordinates attached to them. So while

[3] This is a capability largely enabled through the use of Twitter's web interface, while tweets with precise latitude and longitude coordinates tend to be associated with tweets created on mobile devices.

a macro-level analysis of tweeting across the United States would mean that a tweet simply tagged to 'Lexington, Kentucky' would be a sufficient understanding of a user's location at the time of tweeting, a more micro-scale analysis like some of those presented in this dissertation, requires (nearly) *exact* latitude and longitude coordinates to be attached in order to place people at different locations within the urban context. DOLLY makes this process easier by trying to identify what kind of geotag is associated with each individual tweet – as seen in the 'type' field and the presence of either 'll' for latitude and longitude coordinate, 'p' for place, or 'llp' for both – allowing for a filtering of those tweets that are only tagged to places and without exact latitude and longitude coordinates when necessary.

Figure 1.2: Tweets Downloaded from DOLLY as seen in Microsoft Excel

1	id	u_id	u_description	u_name	u_screen_name	u_lang	u_statuses_count	u_followers_count	u_friends_count	latitude	longitude
2	5.7215E+17	475116753	Retired Army living th	LesHailYes	LesHailYes	en	1958	306	285	37.748873	-87.131843
3	5.7215E+17	144591570	Owner/operator DB's	Dave Baldridge	davidhbaldridge	en	6938	406	643	38.039489	-84.500556
4	5.7216E+17	295881132		erin faith carey.	erinfaithcarey	en	11890	381	355	30.878607	-93.290378
5	5.7217E+17	206245579	Co-host of @CatsTalk	Vinny Hardy	VinnyHardy	en	20122	1312	1895	38.039444	-84.500278
6	5.7219E+17	491889235	sometimes i wish i co	bruke â˜¼	BrookeHinson7	en	3531	431	395	38.652948	-83.961716
7	5.7218E+17	813543086	Class of 2013 @ GCHS	David â€	24david_lewis	en	6235	304	515	38.704363	-84.95882
8	5.7219E+17	20104248	(Iowan) #Nola Transp	Andy Cunninghar	CunninghamWDSU	en	5881	5395	1696	29.926367	-90.076844
9	5.7219E+17	27877541	A proud Veteran,2nd/	Phil Columbia	PhilColumbia	en	8821	1233	1916	38.139085	-85.594938
10	5.722E+17	327690935	If it's worth fighting fc	Courtney Rowlan	CourtneeeLane	en	5765	490	259	38.039889	-84.501385
11	5.722E+17	34744730	University of Kentuck	UK Wildcatsï¼	UKBBTL	en	93600	1490	1863	26.12069	-80.1313
12	5.7221E+17	65397677		Marty Donlin	alrightguy31	en	3980	506	1748	39.006843	-84.617427
13	5.7209E+17	243364216	University of Kentuck	Wilder Treadway	Wilder_Treadway	en	17471	1088	547	42.316716	-72.638313
14	5.7211E+17	3.048E+09	AVA VOLLEYBALL #5	Jaylyn Giesler	GieslerJaylyn	en	21	59	194	38.308628	-86.941884
15	5.7212E+17	144591570	Owner/operator DB's	Dave Baldridge	davidhbaldridge	en	6936	406	643	37.968828	-84.468834
16	5.7213E+17	34791820	Christian, Husband, Fa	Michael Clark	UKDad21	en	6873	796	1969	38.215445	-85.349127
17	5.7221E+17	270924442	Dont follow me... Im I	Courtney Banks	MaddiMommy	en	5815	372	648	30.485683	-81.588046
18	5.7222E+17	2.489E+09	I am the punishment	dill	RawlsDillon	en	3123	160	288	38.129438	-85.877054
19	5.7223E+17	397524943	Chief Meteorologist -	Â©hief â€½ï¸, e	Dthixton10	en	26142	1384	1995	38.077308	-85.653764
20	5.7227E+17	1.12E+09	, CARD FAN ,social ec	Tony	R3L3NTLESS050	en	1711	32	148	38.142703	-85.891768
21	5.7233E+17	15436449	I'm a blast in a glass!	Gaven Blake Oak	gavenoakley	en	5881	708	1268	36.346727	-119.335862
22	5.7244E+17	359619850	I'm a nanny/lifeguard,	Christie Hollis	minasmommyluvuk	en	5433	473	1123	36.038048	-86.728171
23	5.7243E+17	2.411E+09	Lotta love in the hous	WHBlakemore	Woodyman502	en	1484	133	1994	37.8222935	-85.7681905
24	5.7244E+17	2.269E+09	I breed and race stanc	Tara Spach	racing_fillies	en	13186	515	555	38.030266	-84.507996
25	5.7191E+17	2.157E+09		Sean Stiles	Freshstiles10	en	388	16	66	38.326595	-85.527133
26	5.7192E+17	236136464	Brandon nickname BP	BP	oneandonlybp	en	22555	609	598	37.960134	-84.478906

1	type	place_type	area	c_country	c_state	c_county	c_geoid	created_at	hashtags	text
2	llp	city	52100291.3	USA	Kentucky	59	1400000US2105900	########	BBN	@KySportsRadio @ShannonTheDude #BBN @Wildcatcarrie http://t.c
3	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	########	BBN	We do Senior day the best!!! #BBN
4	llp	city		USA	Louisiana	11	1400000US2201196	########	BBN	â€œ@BigBlue9ation: 1 | 41 | 2 | 12 #BBN http://t.co/oxMZl365Oaâ
5	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	########	BBN seeblue	Bowman, Wildcat Statue #BBN #seeblue @ Memorial Coliseum https:
6	llp	admin	1.9729E+11	USA	Kentucky	161	1400000US2116196	########	bbn	â€œ@Andrew_S_Day: RT if you love you're 29-0 Kentucky Wildcats!!
7	llp	admin	1.9729E+11	USA	Kentucky	77	1400000US2107796	########	goladycats b	The #13 UK Ladycats are just killin' it right now against #2 Gamecocks
8	llp	city	908611482	USA	Louisiana	71	1400000US2207100	########	GoBigBlue W	Congrats to UK Women @UKHoopCats on beating No. 2 South Carolii
9	llp	city	14927453.9	USA	Kentucky	111	1400000US2111101	########	BBN	Great game Lady Cats huge win on Senior nite #BBN
10	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	########	BBN	I really do love my school ðŸ˜ #BBN
11	llp	city	93773832.6	USA	Florida	11	1400000US1201104	########	Kentucky BB	RT @JtJtdozier1: #Kentucky is really starting to catch my eyes, getting
12	llp	city	26580575.6	USA	Kentucky	15	1400000US2101507	########	BBN StoopsT	@JtJtdozier1 come be a part of something special in Lexington #BBN
13	llp	city	92494465.6	USA	Massachuset	15	1400000US2501582	########	BBN BeatThe	Would do anything to be at a sold-out Memorial today for @UKHoop
14	llp	admin	1.2583E+11	USA	Indiana	37	1400000US1803795	########	BBN	Counting down for March MadnessðŸ € #BBN
15	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	########	bbn	A Big Yahtzee on a Sunday. #bbn
16	llp	admin	1.9729E+11	USA	Kentucky	211	1400000US2121104	########	BBN	@douggottlieb with some high praise for WCS. Tabbed him as his PO'
17	llp	neighborhood		USA	Florida	31	1400000US1203101	########	BBN	Can't stop, won't stop. #BBN
18	llp	city	21692912.3	USA	Kentucky	111	1400000US2111101	########	bbn	#bbn http://t.co/F6rpqFzOk3
19	llp	city		USA	Kentucky	29	1400000US2102902	########	bbn	â‰¾â‰¾ðŸ †ðŸ †ðŸ † funny.. That night at ur house was fun.. Ha
20	llp	city	21692912.3	USA	Kentucky	111	1400000US2111101	3/2/15 0:20	OffendEvery	Kentucky sucks at basketball #OffendEveryoneIn4Words #CardNation
21	llp	city	80929125	USA	California	107	1400000US0610700	3/2/15 4:54	BBN Kentuck	Hell to the yes! #BBN #Kentucky #ky #Wildcats #NCAA #UK #winnings
22	llp	city		USA	Tennessee	37	1400000US4703701	########	bbn sorrynot	No lie.. I be jammin when our UK commercial comes with the nelly lo
23	p	admin	1.9729E+11	USA	Kentucky	93	1400000US2109398	########	BBN	@therealjaylin11 One less required. #BBN!
24	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	########	stillhateLaet	@juliamarie82 @KySportsRadio I witnessed that "shot" in person. #
25	llp	city		USA	Kentucky	185	1400000US2118503	3/1/15 0:43	BBN	Relaxing and enjoying the Cats being 29-0 with only needing 11 more
26	llp	admin	1.9729E+11	USA	Kentucky	67	1400000US2106700	3/1/15 1:30	BBN	@GarrettFinch106 follow back #BBN

Source: Author's Screenshot

And while DOLLY has already indexed over 11 billion tweets, it is not the sheer size of this data that makes it so useful for geographical research. Rather, it is that such size enables the potential for looking closer at a range of different aspects of each individual data point, allowing for investigations into individuals over time, or to allow for different aggregations of individuals based on common characteristics discerned from their data shadows, or even to simply go back in time to collect data on a given topic months after it had already fallen off the radar. The size of this 'big data' is in the end most meaningful in that it allows for much more precise and targeted 'small datasets' to be analyzed through a variety of quantitative and qualitative methods, as is done throughout the papers that constitute this dissertation. The possibilities for such analyses are explored in each of the three papers of this.

Though the exact methods of analysis differ for each of the three papers that make up this dissertation, there are a few commonalities throughout. The most common strategy throughout the three papers is that of aggregating individual tweets to larger areas using hexagonal bins (cf. Scott 1985 and Carr et al 1992 for a justification of the use of hexagonal binning). This approach simultaneously addresses a range of interconnected problems common in many social media mapping projects. First, simply aggregating individual data points into larger areal units represents an initial fix to the problem of 'overplotting', where densely concentrated individual dots are overlaid on one another to the point of no longer being able to distinguish between different kinds of concentrations (cf. Poorthuis and Zook 2015; Poorthuis et al, forthcoming). Second, the use of hexagonal bins addresses the long-standing issue of the modifiable areal unit problem by utilizing uniformly sized cells. And while hexagonal binning doesn't *eliminate* the modifiable areal unit problem – i.e., the hexagonal bins could theoretically be of any size, thus affecting the distribution of points at different scales – it does minimize some of the arbitrariness associated with census-defined areal units, and allows the researcher to make more context-appropriate decisions about the scale and shape of the areal units being used (see Chapters 3 and 4 for further explorations of the role of spatial scale in shaping these analyses). Finally, the use of hexagonal binning allows us to address the persistent issue of the lack of normalization applied to the data by utilizing somewhat more advanced spatial statistical approaches. In particular, each of the three papers in this dissertation relies heavily on the calculation of odds ratios, also commonly known as the location quotient within economic geography and regional economics, in order to put each individual point in relation to both its spatially-proximate neighbors within a given subset of data, as well as within the broader context of the geography of tweeting activity. By utilizing both a given dataset of interest and a random sample of tweeting activity, the application of the odds ratio allows for an understanding of how tweeting in a given subset compares to overall patterns. The basic formula for the odds ratio is below:

$$OR = \frac{p_i / p}{r_i / r}$$

Where p_i is the number of tweets in hexagon i related to the phenomenon of interest and p is the sum of all tweets related to the phenomenon; r_i is the number of random tweets in hexagon i and r the sum of all random tweets. This results in a ratio where a value of 1 means that there are exactly as many data points for the phenomenon as one would expect based on the random sample. Beyond the simple odds ratio formula, one can calculate the confidence interval for each hexagonal cell, which takes into account the relative abundance or paucity of observations in a given location. In other words, while two different cells may have an odds ratio equal to 1, if one cell has 100 observations and the other cell has only three, greater confidence can be given to the cell with 100 observations, due to the smaller margin of error in the calculation. Below is the formula for the lower bound of the 99.9% confidence interval for each individual cell:

$$OR_{lower} = e^{\ln(OR_i) - 3.29 * \sqrt{\frac{1}{p_i} + \frac{1}{p} + \frac{1}{r_i} + \frac{1}{r}}}$$

In this case, an odds ratio greater than 1 means that we can say, with 99.9% confidence, that there are more points related to the phenomenon than one should expect, and vice versa for anything under 1. Slightly modified versions of this calculation can be made to adjust the confidence level to, say, the lower bound of the 95% confidence interval, with the same result. Beyond the use of hexagonal binning and the use of the odds ratio measure, each of the three papers shares a mixed methods approach that combines this kind of spatial analysis with methods more common to the critical GIScience tradition, such as qualitative analysis of textual content and a focus on more descriptive statistics (cf. Lawson 1995).

There are, however, a number of differences in the methods employed for each paper. As such, key differences and unique elements of each methodology are discussed in turn for each of the three papers. Chapter 2's analysis begins by collecting data associated with a series of keywords related to the shooting death of Michael Brown in Ferguson, Missouri in August 2014. After utilizing the aforementioned methods above for identifying unique concentrations of tweeting about Ferguson, the paper turns to identifying those tweets outside of the US, which were a significant minority in relation to the total dataset, despite being something of a magnet for media attention, as the paper discusses. In order to identify the connections between these tweets from outside the US to the local context, the tweet history for each of the unique users tweeting from outside the US were collected, identifying those users who had not only tweeted from the US before, but had actually tweeted from within the St. Louis metropolitan area.

Chapter 3 utilizes many of the same methods identified previously, focusing on the landfall of Hurricane Sandy in October 2013, although it introduces some additional elements in that it combines the analysis of tweets through the odds ratio measure with ancillary data, both from FEMA's high-impact designation in relation to Hurricane Sandy, as well as RITA data on domestic flights in the United States. In addition to the use of these non-social media datasets for providing context, Chapter 3 also explores in more depth the potentials of qualitative analysis, looking at both tweets about the 57th Street crane during Hurricane Sandy, as well as intra-metropolitan differences in tweeting themes in Los Angeles, far away from the epicenter of the hurricane.

Finally, Chapter 4 diverges from the earlier chapters by utilizing an alternative means of data collection. Rather than attempting to collect tweets based on the usage of keywords about socio-spatial inequality, data was collected based on longer histories of tweeting activity in different areas of the city in order to produce parallel datasets of those who can be identified as 'belonging' to different neighborhoods, so as to examine the relationships between these different neighborhoods and popular socio-spatial imaginaries tied to them. This particular method of collecting data allows for different types of contextualization through the odds ratio measure, comparing parallel groups rather than a primary dataset to a random sample, while also avoiding the issue of collecting 'false positives' in the form of tweets that use a given keyword, albeit in a different way than was intended during the data collection.

V. Overview of the Dissertation

This dissertation is composed of three separate papers that share a broad focus on articulating an alternative conceptual and methodological approach to the study of geotagged social media data. While each paper deals with a distinct case study or issue, the papers together move sequentially through the different elements of Brinberg and McGrath's (1985) triad of conceptual, methodological and substantive knowledges.

A. Paper #1: Spatialities of Data

The first paper in this dissertation, entitled "Spatialities of Data", resides largely within the conceptual arena. This paper seeks to answer the question, How have geographers and big data researchers conceptualized the spatiality of big data, and how do these conceptualizations constrain or enable the analysis of geotagged social media data? By describing in more depth the 'spatial ontology of the geotag' and how it has been operationalized in a series of prominent social media mapping projects undertaken by the data journalist Simon Rogers, this paper sets the groundwork for the critique at the heart of this dissertation. That is, this paper argues that the overly simplistic conceptualizations of space operationalized in prominent social media mapping projects foreclose significant potentials for making more substantive insights into the world using this kind of geotagged social media data, which in turn has the effect of dissuading critical scholars from making substantive engagements with this data in their research. The paper then uses a pair of case studies – one of tweeting about the events in Ferguson, Missouri following the police shooting of an African-American teenager in August 2014, one focused on racialized inequalities in Louisville, Kentucky – to demonstrate how a conceptualization of space drawn from relational socio-spatial theory, combined with more context-sensitive methods of data collection, statistical analysis and cartographic representation, can allow for substantively different understandings of the social issues at hand when mapping social media data.

B. Paper #2: Mapping the Data Shadows of Hurricane Sandy

The second paper in the dissertation, "Mapping the Data Shadows of Hurricane Sandy", focuses on taking the conceptual insights of the first paper, and of a relational or multidimensional socio-spatial theory more broadly, and having these concepts inform the methodological approach for analyzing geotagged social media data. In asking how can existing frameworks for conceptualizing the multidimensionality of socio-spatial relations be usefully applied to the analysis of big data, this paper serves as a proof-of-concept for this dissertation's central thesis. That is, this paper demonstrates the fundamental compatibility of these concepts, methods and data sources, despite the various critiques of big data as being somehow incommensurable with these longer-standing approaches within geographic research. More specifically, this paper applies the TPSN framework developed by Jessop et al (2008) to the empirical case of tweeting in the wake of Hurricane Sandy in October 2013. By demonstrating how different dimensions of spatiality can be enacted in an analysis of geotagged social media data, this paper shows how an analysis of events like Hurricane Sandy are reflected in a variety of ways through the prism of social media data, and have more complex geographies than simple concentrations in a given place.

C. Paper #3: Social Media and the City

The third and final paper in this dissertation, "Social Media and the City", takes the broader conceptual and methodological insights from the first and second papers, and builds on them by asking: how can relational socio-spatial theory and critical/qualitative GIScience be applied to big data in order to produce alternative ways of imagining urban spaces and places and the inequalities between them?

Rather than simply pointing to the possibilities of using this data in a more nuanced and situated manner, this paper offers a substantive intervention into a 'real world', 'on-the-ground' issue, in this case questions of urban socio-spatial imaginaries and inequalities in Louisville, Kentucky. This paper starts by questioning the dominant popular socio-spatial imaginary of the '9th Street Divide', a colloquial naming of the division between the city's predominantly poor and black West End neighborhoods and the rest of the city's more affluent and largely white population. Rather than reinforcing segregation in an absolute way, this paper argues that geotagged social media data offers an opportunity to develop more nuanced understandings of segregation by understanding people's everyday mobilities over time, where different kinds of people move through the city and how they utilize it in different ways.

By classifying a number of local Twitter users as 'belonging' to either the West End or a comparable area of the city's East End, we can visualize these individuals' movements over time, understanding with more precision how processes of inequality and segregation operate in people's everyday lives. Through such an analysis, this paper argues that while the 9th Street Divide remains a salient way of summarizing the persistence of racial inequality in Louisville, it actually conceals the fact that it is the predominantly wealthy and white East End residents who are more spatially segregated within the city, while the long-term effects of racial residential segregation has actually meant that West End residents have become incredibly spatially mobile within the city. Ultimately, this analysis points towards both an alternative empirical reality of how inequality operates within the city, as well as to a broader understanding of the West End as spatially extensive and fluid, rather than isolated or fundamentally separate and apart from the rest of the city as is commonly remarked.

VI. References

Allen, John, and Allan Cochrane. 2007. "Beyond the Territorial Fix: Regional Assemblages, Politics and Power." *Regional Studies* 41 (9): 1161–1175.

Amin, Ash. 2002. "Spatialities of Globalisation." *Environment and Planning A* 34 (3): 385–399.

Amin, Ash. 2007. "Re-thinking the Urban Social." *City* 11 (1): 100–114.

Amin, Ash, and Nigel Thrift. 2002. *Cities: Reimagining the Urban*. Polity.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine* 15(7).

Barnes, Trevor J. 2013. "Big Data, Little History." *Dialogues in Human Geography* 3 (3): 297–302.

Barnes, Trevor J., and Matthew W. Wilson. 2014. "Big Data, Social Physics, and Spatial Analysis: The Early Years." *Big Data & Society* 1 (1).

Batty, Michael. 2013. *The New Science of Cities*. MIT Press.

Bettencourt, Luis, and Geoffrey West. 2010. "A Unified Theory of Urban Living." *Nature* 467 (7318): 912–913.

Bowker, Geoffrey C. 2014. "The Theory/Data Thing." *International Journal of Communication* 8: 1795–1799.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–679.

Brenner, Neil. 2013. "Theses on Urbanization." *Public Culture* 25 (1): 85–114.

Brenner, Neil, David J. Madden, and David Wachsmuth. 2011. "Assemblage Urbanism and the Challenges of Critical Urban Theory." *City* 15 (2): 225–240.

Brenner, Neil, and Christian Schmid. 2014. "The 'Urban Age' in Question." *International Journal of Urban and Regional Research* 38 (3): 731–755.

Brinberg, David, and Joseph E. McGrath. 1985. *Validity and the Research Process.* Sage.

Carr, Daniel B., Anthony R. Olsen, and Denis White. 1992. "Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data." *Cartography and Geographic Information Systems* 19 (4): 228–236.

Chen, Xiang, and Xining Yang. 2014. "Does Food Environment Influence Food Choices? A Geographical Analysis through 'tweets.'" *Applied Geography* 51: 82–89.

Cope, Meghan, and Sarah Elwood. 2009. *Qualitative GIS: A Mixed Methods Approach*. Sage.

Crampton, Jeremy W. 2010. *Mapping: A Critical Introduction to Cartography and GIS*. John Wiley & Sons.

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb." *Cartography and Geographic Information Science* 40 (2): 130–139.

Cranshaw, Justin, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. "The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a

City." In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, June, pp. 58–65.

Crawford, Kate. 2013. "The Hidden Biases in Big Data." *Harvard Business Review*. 1 April. Available from: http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html.

Crutcher, Michael, and Matthew Zook. 2009. "Placemarks and Waterlines: Racialized Cyberscapes in Post-Katrina Google Earth." *Geoforum* 40 (4): 523–534.

Curry, Michael R. 1997. "The Digital Individual and the Private Realm." *Annals of the Association of American Geographers* 87 (4): 681–689.

Dialogues in Human Geography. 2013. Forum on "Geography and the Future of Big Data, Big Data and the Future of Geography". 3 (3).

Dodge, Martin, and Rob Kitchin. 2013. "Crowdsourced Cartography: Mapping Experience and Knowledge." *Environment and Planning A* 45 (1): 19–36.

Elwood, Sarah A. 2002. "GIS Use in Community Planning: A Multidimensional Analysis of Empowerment." *Environment and Planning A* 34 (5): 905–922.

Elwood, Sarah. 2006. "Beyond Cooptation or Resistance: Urban Spatial Politics, Community Organizations, and GIS-Based Spatial Narratives." *Annals of the Association of American Geographers* 96 (2): 323–341.

Elwood, Sarah. 2010. "Thinking Outside the Box: Engaging Critical Geographic Information Systems Theory, Practice and Politics in Human Geography." *Geography Compass* 4 (1): 45–60.

Elwood, Sarah, Michael F. Goodchild, and Daniel Z. Sui. 2012. "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice." *Annals of the Association of American Geographers* 102 (3): 571–590.

Elwood, Sarah, and Helga Leitner. 2003. "GIS and Spatial Knowledge Production for Neighborhood Revitalization: Negotiating State Priorities and Neighborhood Visions." *Journal of Urban Affairs* 25 (2): 139–157.

Elwood, Sarah, and Agnieszka Leszczynski. 2011. "Privacy, Reconsidered: New Representations, Data Practices, and the Geoweb." *Geoforum* 42 (1): 6–15.

Elwood, Sarah, and Agnieszka Leszczynski. 2013. "New Spatial Media, New Knowledge Politics." *Transactions of the Institute of British Geographers* 38 (4): 544–559.

Feick, Rob, and Colin Robertson. Forthcoming. "A Multi-Scale Approach to Exploring Urban Places in Geotagged Photographs." *Computers, Environment and Urban Systems*.

Ghose, Rhina. 2001. "Use of Information Technology for Community Empowerment: Transforming Geographic Information Systems into Community Information Systems." *Transactions in GIS* 5 (2): 141–163.

Ghosh, Debarchana (Debs), and Rajarshi 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.

Goodchild, Michael. 2006. "GIScience Ten Years After Ground Truth." *Transactions in GIS* 10 (5): 687–692.

Goodchild, Michael. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69 (4): 211–221.

Goodchild, Michael. 2009. "NeoGeography and the Nature of Geographic Expertise." *Journal of Location Based Services* 3 (2): 82–96.

Goss, Jon. 1995. "'We Know Who You Are and We Know Where You Live': The Instrumental Rationality of Geodemographic Systems." *Economic Geography* 71 (2): 171–198.

Graham, Mark. 2010. "Neogeography and the Palimpsests of Place: Web 2.0 and the Construction of a Virtual Earth." *Tijdschrift Voor Economische En Sociale Geografie* 101 (4): 422–436.

Graham, Mark, Scott A. Hale, and Devin Gaffney. 2014. "Where in the World Are You? Geolocation and Language Identification in Twitter." *The Professional Geographer* 66 (4): 568–578.

Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. "Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty." *Annals of the Association of American Geographers* 104 (4): 746–764.

Graham, Mark, and Taylor Shelton. 2013. "Geography and the Future of Big Data, Big Data and the Future of Geography." *Dialogues in Human Geography* 3 (3): 255–261.

Graham, Mark, and Matthew Zook. 2011. "Visualizing Global Cyberscapes: Mapping User-Generated Placemarks." *Journal of Urban Technology* 18 (1): 115–132.

Graham, Mark, and Matthew Zook. 2013. "Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web." *Environment and Planning A* 45 (1): 77–99.

Graham, Stephen, and Simon Marvin. 2001. *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. Routledge.

Haklay, Muki. 2013. "Neogeography and the Delusion of Democratisation." *Environment and Planning A* 45 (1): 55–69.

Han, Su Yeon, Ming-Hsiang Tsou, and Keith C. Clarke. 2015. "Do Global Cities Enable Global Views? Using Twitter to Quantify the Level of Geographical Awareness of U.S. Cities." *PLoS ONE* 10 (7).

Hemsley, Jeff, and Josef Eckert. 2014. "Examining the Role of 'Place' in Twitter Networks through the Lens of Contentious Politics." In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, pp. 1844–1853.

Jessop, Bob, Neil Brenner, and Martin Jones. 2008. "Theorizing Sociospatial Relations." *Environment and Planning D: Society and Space* 26 (3): 389–401.

Kelley, Matthew James. 2013. "The Emergent Urban Imaginaries of Geosocial Media." *GeoJournal* 78 (1): 181–203.

Kitchin, Rob. 2013. "Big Data and Human Geography: Opportunities, Challenges and Risks." *Dialogues in Human Geography* 3 (3): 262–267.

Kitchin, Rob. 2014a. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.

Kitchin, Rob. 2014b. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1).

Kitchin, Rob, and Tracey P. Lauriault. 2015. "Small Data in the Era of Big Data." *GeoJournal* 80 (4): 463–475.

Knigge, LaDona, and Meghan Cope. 2006. "Grounded Visualization: Integrating the Analysis of Qualitative and Quantitative Data through Grounded Theory and Visualization." *Environment and Planning A* 38 (11): 2021–2037.

Kwan, Mei-Po. 2002. "Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research." *Annals of the Association of American Geographers* 92 (4): 645–661.

Kwan, Mei-Po. 2013. "Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility." *Annals of the Association of American Geographers* 103 (5): 1078–1086.

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity, and Variety". META Group. Available from: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lawson, Victoria. 1995. "The Politics of Difference: Examining the Quantitative/Qualitative Dualism in Post-Structuralist Feminist Research." *The Professional Geographer* 47 (4): 449–457.

Leszczynski, Agnieszka. 2012. "Situating the Geoweb in Political Economy." *Progress in Human Geography* 36 (1): 72–89.

Leszczynski, Agnieszka. Forthcoming. “Spatial Big Data and Anxieties of Control.” *Environment and Planning D: Society and Space.*

Lin, Jie, and Robert G. Cromley. 2015. “Evaluating Geo-Located Twitter Data as a Control Layer for Areal Interpolation of Population.” *Applied Geography* 58: 41–47.

Longley, Paul A., Muhammad Adnan, and Guy Lansley. 2015. “The Geotemporal Demographics of Twitter Usage.” *Environment and Planning A* 47 (2): 465–484.

MacLeod, Gordon, and Martin Jones. 2007. “Territorial, Scalar, Networked, Connected: In What Sense a ‘Regional World’?” *Regional Studies* 41 (9): 1177–1191.

Manyinka, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. “Big Data: The next Frontier for Innovation, Competition, and Productivity”. McKinsey Global Institute. Available from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation/

Massey, Doreen. 1991. “A Global Sense of Place.” *Marxism Today* 35 (6): 24–29.

McCann, Eugene, and Kevin Ward. 2010. “Relationality/territoriality: Toward a Conceptualization of Cities in the World.” *Geoforum* 41 (2): 175–184.

McFarlane, Colin. 2011. “Assemblage and Critical Urbanism.” *City* 15 (2): 204–224.

Miller, Greg. 2011. “Social Scientists Wade Into the Tweet Stream.” *Science* 333 (6051): 1814–1815.

Miller, Harvey, and Michael Goodchild. 2015. “Data-driven geography.” *GeoJournal* 80 (4): 449–461.

Miller, Roger P. 1995. “Beyond Method, Beyond Ethics: Integrating Social Theory into GIS and GIS into Social Theory.” *Cartography and Geographic Information Systems* 22 (1): 98–103.

Noulas, Anastasios, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2012. “A Tale of Many Cities: Universal Patterns in Human Urban Mobility.” *PLoS ONE* 7 (5).

Pavlovskaya, Marianna. 2002. “Mapping Urban Change and Changing GIS: Other Views of Economic Restructuring.” *Gender, Place & Culture* 9 (3): 281–289.

Pavlovskaya, Marianna. 2006. “Theorizing with GIS: A Tool for Critical Geographies?” *Environment and Planning A* 38 (11): 2003–2020.

Pentland, Alex. 2014. *Social Physics: How Good Ideas Spread-The Lessons from a New Science*. Penguin.

Pickles, John. 1995a. *Ground Truth: The Social Implications of Geographic Information Systems*. Guilford Press.

Pickles, John. 1995b. "Representations in an Electronic Age: Geography, GIS, and Democracy." In *Ground Truth: The Social Implications of Geographic Information Systems*, edited by John Pickles. Guilford Press. pp. 1-30.

Poorthuis, Ate, and Matthew Zook. 2015. "Small Stories in Big Data: Gaining Insights From Large Spatial Point Pattern Datasets." *Cityscape* 17 (1): 151–160.

Poorthuis, Ate, Matthew Zook, Taylor Shelton, Mark Graham, and Monica Stephens. Forthcoming. "Using Geotagged Digital Social Data in Geographic Research." In *Key Methods in Geography (3rd Edition)*, edited by Nick Clifford, Shaun French, Meghan Cope, and Thomas Gillespie. Sage.

Press, Gil. 2013. "$16.1 Billion Big Data Market: 2014 Predictions from IDC and IIA". *Forbes.* 12 December. Available from: http://www.forbes.com/sites/gilpress/2013/12/12/16-1-billion-big-data-market-2014-predictions-from-idc-and-iia/

Preston, Bryan, and Matthew W. Wilson. 2014. "Practicing GIS as Mixed Method: Affordances and Limitations in an Urban Gardening Study." *Annals of the Association of American Geographers* 104 (3): 510–529.

Rundstrom, Robert A. 1995. "GIS, Indigenous Peoples, and Epistemological Diversity." *Cartography and Geographic Information Systems* 22 (1): 45–57.

Salesses, Philip, Katja Schechtner, and César A. Hidalgo. 2013. "The Collaborative Image of The City: Mapping the Inequality of Urban Perception." *PLoS ONE* 8 (7).

Savage, Mike, and Roger Burrows. 2007. "The Coming Crisis of Empirical Sociology." *Sociology* 41 (5): 885–899.

Schuurman, Nadine. 2000. "Trouble in the Heartland: GIS and Its Critics in the 1990s." *Progress in Human Geography* 24 (4): 569–590.

Scott, David W. 1985. "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions." *The Annals of Statistics* 13 (3): 1024–1040.

Sheller, Mimi, and John Urry. 2006. "The New Mobilities Paradigm." *Environment and Planning A* 38 (2): 207–226.

Shelton, Taylor, Matthew Zook, and Mark Graham. 2012. "The Technology of Religion: Mapping Religious Cyberscapes." *The Professional Geographer* 64 (4): 602–617.

Sheppard, Eric. 1995. "GIS and Society: Towards a Research Agenda." *Cartography and Geographic Information Science* 22 (1): 5–16.

Sheppard, Eric. 2001. "Quantitative Geography: Representations, Practices, and Possibilities." *Environment and Planning D: Society and Space* 19 (5): 535–554.

Smith, Neil. 1992. "History and Philosophy of Geography: Real Wars, Theory Wars." *Progress in Human Geography* 16 (2): 257–271.

Stefanidis, Anthony, Amy Cotnoir, Arie Croitoru, Andrew Crooks, Matthew Rice, and Jacek Radzikowski. 2013. "Demarcating New Boundaries: Mapping Virtual Polycentric Communities through Social Media Content." *Cartography and Geographic Information Science* 40 (2): 116–129.

Taylor, Peter J. 1990. "GKS." *Political Geography Quarterly* 9 (3): 211–212.

Thatcher, Jim. 2014. "Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data." *International Journal of Communication* 8: 1765–1783.

Tsou, Ming-Hsiang, Jiue-An Yang, Daniel Lusher, Su Han, Brian Spitzberg, Jean Mark Gawron, Dipak Gupta, and Li An. 2013. "Mapping Social Activities and Concepts with Social Media (Twitter) and Web Search Engines (Yahoo and Bing): A Case Study in 2012 US Presidential Election." *Cartography and Geographic Information Science* 40 (4): 337–348.

Wall, Melissa, and Treepon Kirdnark. 2012. "Online Maps and Minorities: Geotagging Thailand's Muslims." *New Media & Society* 14 (4): 701–716.

Warf, Barney, and Daniel Sui. 2010. "From GIS to Neogeography: Ontological Implications and Theories of Truth." *Annals of GIS* 16 (4): 197–209.

Watkins, Derek. 2012. "Digital Facets of Place: Flickr's Mappings of the U.S.-Mexico Borderlands". Unpublished M.A. Thesis, University of Oregon Department of Geography.

Widener, Michael J., and Wenwen Li. 2014. "Using Geolocated Twitter Data to Monitor the Prevalence of Healthy and Unhealthy Food References across the US." *Applied Geography* 54: 189–197.

Wilson, Matthew W. 2015. "Morgan Freeman Is Dead and Other Big Data Stories." *Cultural Geographies* 22 (2): 345–349.

Wyly, Elvin. 2009. "Strategic Positivism." *The Professional Geographer* 61 (3): 310–322.

Wyly, Elvin. 2011. "Positively Radical." *International Journal of Urban and Regional Research* 35 (5): 889–912.

Wyly, Elvin. 2014. "The New Quantitative Revolution." *Dialogues in Human Geography* 4 (1): 26–38.

Xu, Chen, David W. Wong, and Chaowei Yang. 2013. “Evaluating the ‘Geographical Awareness’ of Individuals: An Exploratory Analysis of Twitter Data.” *Cartography and Geographic Information Science* 40 (2): 103–115.

Yapa, Lakshman. 1998. “Why GIS Needs Postmodern Social Theory, and Vice Versa.” In *Modern Cartography Series, Volume 3: Policy Issues in Modern Cartography*, edited by Fraser Taylor. Academic Press. pp. 249–69.

Zhai, Shixiao, Xiaolin Xu, Lanrong Yang, Min Zhou, Lu Zhang, and Bingkui Qiu. 2015. “Mapping the Popularity of Urban Restaurants Using Social Media Data.” *Applied Geography* 63: 113–120.

Zook, Matthew, and Mark Graham. 2010. “Featured Graphic: The Virtual ‘Bible Belt.’” *Environment and Planning A* 42 (4): 763–764.

Zook, Matthew, and Ate Poorthuis. 2014. “Offline Brews and Online Views: Exploring the Geography of Beer Tweets.” In *The Geography of Beer*, edited by Mark Patterson and Nancy Hoalst-Pullen. Springer. pp. 201-209.

Chapter 2
Spatialities of Data:
Mapping Social Media ‘Beyond the Geotag’

For submission to *GeoJournal*

I. Introduction

Throughout all corners of society, data – as an organizing idea and set of practices – is becoming increasingly central to the ways that individuals and organizations think about the world and their actions within it. While data can take many forms – big or small, open or proprietary, digital or analog, volunteered or captured – this attraction to data has largely been driven by the somewhat recent emergence of so-called 'big data' and the associated "widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy" (boyd and Crawford 2012: 663). Such a sentiment is perhaps best illustrated by Chris Anderson, the former editor of *Wired*, in his now-infamous celebration of big data, when he wrote that "[w]ith enough data, the numbers speak for themselves" (Anderson 2008).

Of particular note within the umbrella of big data is the wealth of data generated via social media platforms like Twitter, Facebook, Foursquare and Instagram, among others. Because of its relative accessibility compared to other proprietary data streams controlled by government agencies or corporate actors – not to mention the general popularity and near-ubiquity of social media throughout much of society – this data has become an increasingly popular starting point for those wishing to undertake social research utilizing big data. This has been especially true for both those 'social physicists' seeking to make the jump from using computational tools to study physical systems to studying social processes, as well as for spatially-oriented social scientists, who have taken advantage of the fact that a significant amount of this data – even if only a small proportion of the total – includes explicit geographic references in the form of a latitude and longitude coordinate pair, or 'geotag'.

Using a variety of techniques, these geographically-oriented studies have focused on how geotagged social media data can be used for everything from identifying the relationship between the 'happiness' of different places and overall quality-of-life indicators (Mitchell et al 2013), understanding how individual food consumption habits are shaped by the surrounding food environment (Chen and Yang 2014; Widener and Li 2014), locating the social epicenter of natural disasters (Crooks et al 2013) and how the digital reflections of disaster response are shaped by offline inequalities (Crutcher and Zook 2009), understanding how people move through space from the urban to the global scale (Hawelka et al 2014; Fischer 2010), predicting levels of unemployment (Llorente et al 2014), understanding the connection between place-based cultural identities and their reflections in digital spaces (Graham and Zook 2011, 2013; Shelton et al 2012) and understanding how these new social networks reconfigure the spatialities of interpersonal relationships (Leetaru et al 2013; Takhteyev et al 2012).

But as these studies have proliferated, this nascent, if nebulous, subfield of social media mapping has also come under fire from critical scholars for promoting a kind of 'speedy pseudopositivism' associated with a neoliberalizing 'new quantitative revolution' (Wyly 2014). As Wyly argues, "[big data] can be ruthlessly ahistorical", providing little in the way of meaningful insight, but instead producing only "a quickly expanding, shallow view of the vast horizontal landscape of the desert of the present real…accomplishing new kinds of devalorization of past generations of human knowledge" (Wyly 2014: 28). But in seemingly dismissing all of this data out of hand, Wyly fails to explore the possible synergies between the analysis of big data – and social media data, in particular – and a variety of post- and non-positivist epistemologies, in accordance with his earlier call to rethink the mid-20th century geography's contingent and historically-specific connection between quantitative methods, positivist epistemology and reactionary politics that has since checkered much of the discipline's engagement with such forms of quantitative analysis (Wyly 2009, 2011). That is, while the origins of this data within profit-maximizing corporate organizations and their occasionally naïve and/or malicious use are deserving of our critical attention, these facts should not dissuade us entirely from pursuing alternative forms of engagement with this data. Indeed, the fact that geography's disciplinary history has already been marked by the necessity of engaging in such quantitative scholarship provides something of a roadmap for how to produce more constructive engagements with big data moving forward (Sheppard 2001; Barnes 2013; Graham and Shelton 2013).

As Kitchin has argued, "it *is* possible to think of new epistemologies that do not dismiss or reject Big Data analytics, but rather employ the methodological approach of data-driven science within a different epistemological framing that enables social scientists to draw valuable insights from Big Data that are situated and reflexive" (Kitchin 2014a: 9-10, emphasis added). More specifically, this paper argues that one key point of conflict preventing a productive exchange between the longstanding critical tradition within geography (as represented, at least partially, by Wyly's critique) and those engaged in the project of mapping and analyzing social media data is around the conceptualization of space and spatiality as it applies to this data. All too often, even within some academic circles, questions of how to conceive of space and spatiality are pushed into the background (Massey 1999). Even Kitchin's (2014b) comprehensive deconstruction of 'data' as a conceptual object only goes as far as to argue that data are a geographic phenomena – i.e., shaped by the particular geographic context out of which they emerge – while ultimately avoiding the question of how the spatiality of data might itself be conceptualized.

In the case of analyzing geotagged social media data, this failure to conceptualize space has meant that an often implicit 'spatial ontology of the geotag' has become pervasive in many analyses (Crampton et al 2013). That is, in mapping geotagged social media data, analysts often over-privilege the single pair of latitude and longitude coordinates that are attached to each individual piece of data, "ignoring the multiplicity of ways that space is implicated in the creation of such data" (Crampton et al 2013: 132) by reducing each piece of data to its latitude/longitude coordinate pair. As Crampton et al continue, "a piece of information geotagged to a particular location may not necessarily have been produced in that location, be about that location, or exclude reference to any other geographic locality. Indeed, myriad examples suggest that geotagged content often exhibits a variety of spatial referents apart from the hidden latitude/longitude coordinates attached to it" (Crampton et al 2013: 132). But even if such understandings of space are not articulated explicitly, this implicit conceptualization of space remains crucial in shaping the kind of analysis performed and the conclusions drawn from it. And by failing to attend to a range of social and spatial processes embedded in this kind of data, those mainstream social media mapping projects – not to mention a number of more academically-oriented projects undertaken by non-social scientists – can lead to a range of decontextualized, problematic assertions, as alluded to in earlier critiques of the increasing shift towards and amateurization and privatization of GIScience (Sui 2008; Crampton 2010; Wilson 2015). These kinds of problematic assertions, in turn, only further alienate a range of scholars from critical engaging with the possibilities this data offers for a more grounded and contextualized socio-spatial analysis.

With this in mind, the goals of this paper are two-fold: first, to expand the critique of the spatial ontology of the geotag by outlining the incongruences between the largely implicit conceptions of space within 'mainstream' social media mapping exercises and those more explicit conceptions of geographers; and second, to outline how the integration of relational socio-spatial theory and critical/qualitative GIScience allows for a more geographically-situated analysis of social media data, yielding substantially different understandings of the underlying social processes embedded in such data. Beginning with a case study of how the online reaction to the murder of

African-American teenager Michael Brown in Ferguson, Missouri allows for a rethinking of the geography of social media, and then turning to how this data can allow for a reconsideration of neighborhood-level socio-spatial inequalities in Louisville, Kentucky, this paper points toward a fruitful trading ground between those seeking to utilize these new sources of data for social and spatial research and those more critically-oriented social scientists who have remained skeptical of such data due to its seeming incompatibility with existing epistemological and methodological frameworks.

II. Conceptualizing Space and Spatiality in Social Media Mapping

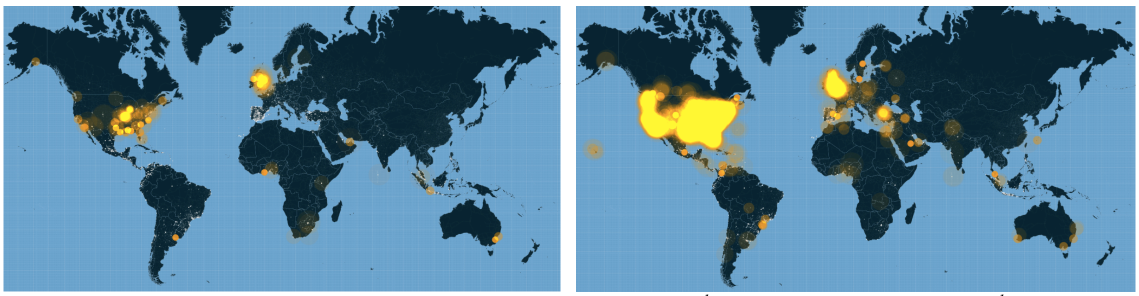
As the broader 'spatial turn' in the social sciences and humanities has taken hold (Warf and Arias 2008), the geographical dimensions of a range of social phenomena have taken center stage. But despite this resurgence of interest in geography and geospatial technologies, the dominant conceptualizations of space mobilized across these disciplines remain tied to long-since superseded Cartesian or Newtonian understandings of space as physical and absolute, an inert plane or container within or on which social relations occur (Curry 1995). As Edward Soja writes, "the term spatial typically evokes the image of something physical and external to the social context and to social action, a part of the 'environment,' a context *for* society - its container - rather than a structure created *by* society" (Soja 1980: 210; emphasis in original). Space in this Cartesian conception pre-exists social relations, has definitive boundaries, is internally coherent, and is tied to particular territorial demarcations, such as the city, the region, or the nation-state. It can, perhaps most importantly for our purposes, be easily mapped because of the definitive nature of its geometry, the latitude and longitude coordinates that organize different spaces in relation to one another.

It is this Cartesian ideal of space as divorced from social relations that underpins the aforementioned spatial ontology of the geotag dominant in many contemporary examples of social media mapping. Arguably the most prominent example of how this spatial ontology is employed in mainstream or 'popular' social media mapping is the portfolio of maps created by data journalist Simon Rogers as part of his previous employment by Twitter. Using prepackaged mapping tools from the mapping start-up CartoDB, Rogers has created a wealth of maps of geotagged tweeting activity on many topics including the World Cup in 2014, the surprise release of a Beyoncé album, the gravely-serious murder of French cartoonists at the satirical magazine *Charlie Hebdo*, and public outcry over police violence in Ferguson, Missouri. Regardless of the topic, Rogers' maps have been routinely distributed throughout a range of popular online media outlets, often with catchy headlines proclaiming the potential insights into the landscape of online social media and society writ-large that one can gain from viewing and interpreting such maps. Rogers' maps are repeatedly described as 'amazing', as well as 'mesmerizing', 'incredible' and 'stirring', sentiments CartoDB proudly trumpets on its own website (CartoDB 2014).

These maps, however, present a substantive problem for those researching the geography of social media data that is belied by their often celebratory reception in the media, a problem owed in large part to their privileging of an absolutist conception of both space and time. That is, Rogers' maps over-privilege the existence of a latitude/longitude coordinate pair – as well as a discrete, identifiable timestamp – attached to each individual point, without looking at the wealth of context that might be drawn out of such data to answer more substantive questions about the phenomena in question. This focus on the absolute is similarly manifest in the assumption that the sheer volume of data – noted in such citable quantitative figures like "3.5 million tweets" – is sufficient to warrant attention and analysis, regardless of what the particular phenomena or its spatial manifestation might be. That is, like a number of other prominent social media mapping projects (cf. Fischer 2014), Rogers' maps equate *more* data with a necessarily improved, *better* understanding of the phenomena at hand, an important corollary to the spatial ontology of the geotag discussed above.

So we can discern from Rogers' maps and the occasional accompanying statistic that *something* important is happening here. However, because the relationships between these individual data points and the places and times they were created in are invisible to us we are unable to discern what exactly it is that we're supposed to be seeing. Even as the maps allow for some understanding of the spatial diffusion of tweets about a given topic over time, these maps inevitably devolve into an undifferentiated flashing blob that resembles little more than a map of population density – or, more accurately, the density of the tweeting population – in the developed world. That the 3.5 million tweets about the grand jury decision in Ferguson in a 24-hour period in late November 2014 end up looking largely indistinguishable from the 14.7 million tweets created about the 2014 Oscars in just a three-and-a-half hour span is indicative of the difficulty associated with interpreting these visualizations. It is this visual effect that has lead the cartographer Kenneth Field to derisively label these maps ‘animated ectoplasm’ (Field 2014).

Figure 2.1: “Tweets mentioning Ferguson” by Simon Rogers



Source: Author’s Screenshot. Map is shown as of August 9th (on left) and August 16th (on right). Full, interactive map available from: http://srogers.cartodb.com/viz/4a5eb582-23ed-11e4-bd6b-0e230854a1cb/embed_map

The expansion of this ‘animated ectoplasm’ across each of Rogers’ maps produces a somewhat ironic understanding of tweeting as an *aspatial* phenomenon. That is, these maps show tweeting to be so pervasive and spatially extensive that its geography is largely unimportant due precisely to its universality. Although CartoDB employee Andrew Hill defends the maps by arguing that “Twitter is making no comment about relative activity from one location to another” (Hill 2015), this is a move he himself describes as ‘slightly deceptive’, perhaps due to the maps never having any accompanying textual explanation. However, the fact that Rogers’ maps have been so widely-shared and the spatial patterns interpreted suggests that there is something going on here beyond simply “communicating the impact and relevance of Twitter in an online and global conversation” (Hill 2015). As of June 2015, Rogers’ original interactive map of Ferguson-related tweeting (see Figure 2.1) has been viewed over 227,000 times, with links to the map appearing on roughly 300 other websites and being shared over 1,400 times via Twitter itself and engaged with over 1,000 times via Facebook. Rogers' follow-up map of reactions to the November 24 grand jury decision not to indict Officer Darren Wilson has been viewed over 3.2 million times, with over 700 other webpages and 3,700 tweets containing links to the map[4].

Among the interpretations of Rogers' initial map of tweeting in reaction to the shooting of Michael Brown in Ferguson, Missouri, one *Washington Post* reporter focused on the flashes occurring in places quite far from the locus of the events, stating, "People are watching from as far away as Fiji and Ghana. That's the world we live in now" (Fung 2014). Others similarly focused on the seemingly complete attention to this event by pointing to how "the impact of the events…can be measured on a global scale" (Capps 2014) or how the massive growth of tweets across the country and the globe could be likened to "a global thermonuclear war…played out on the internet" (Brownlee 2014). In this case, such an astonishment that someone half a world away might be tweeting about Ferguson serves only to reinforce the persistent notion that the growth of the internet has lead to a 'death of distance', whereby such social phenomena are no longer defined by proximity or propinquity, but are fully entangled and uniformly interconnected across great spatial distances (cf. Kirsch 1995 for a critique). Rather than highlighting the particularity of social phenomena in space and the

[4] Statistics via www.sharedcount.com and www.google.com

importance of these offline geographies to their online reflections, Rogers' maps tend to do away with the more complex understandings of space and spatiality developed by geographers over the past two to three decades.

Because this particular subset of data is decontextualized in both space and time as it is presented in this animated map, we are unable to compare these patterns to either more general levels of tweeting in particular places over a longer period of time, or to other topics that might have been trending simultaneously, so as to understand what in particular is unique about the spatial and temporal diffusion of discussions about Ferguson. Casual consumers of these visualizations are unable to determine how the 3.5 million tweets on November 24 reacting to a grand jury's decision not to indict Darren Wilson for Michael Brown's murder are related to tweeting about other topics or to more general levels of tweeting, or how the geography of this tweeting activity compares across such issues. When comparing these 3.5 million tweets to the aforementioned 14.7 million tweets about the 2014 Oscars, the map would seem to show anything but the "engaged American public" trumpeted by NBC News (2014). Instead, these maps reorient the focus onto these events as novel and fleeting, bolstered by the automatic sorting of Twitter's Trending Topics algorithm, which privileges those discussions which grow instantaneously, rather than gradually over time (Lotan 2011; see also Sullivan 2014). So, ultimately, this way of viewing the world through animated Twitter maps promotes an understanding of the social phenomena at hand, in this case the reaction to racialized police brutality, as temporally-specific and spatiallyindiscriminate. But if one is talking about the reactions to – and social experiences of – racism and racialized police violence, these issues are anything but; they are instead incredibly durable over time and, while remarkably pervasive, highly specific and targeted at particular kinds of spaces and places. As Bonilla and Rosa argue:

> "It is thus important to recognize that the reactions to the death of Michael Brown did not spark in a vacuum; they were fueled by accumulated frustrations over previously mediatized moments of injustice and guided by previous digital campaigns. This aggregative effect powerfully positions different instances of racialized brutality not simply as isolated contemporary phenomena but as long-standing systematic forms of violence" (Bonilla and Rosa 2015: 10)

Even if the intent of Rogers' map is only to reinforce Twitter's positionality as a key medium through which we perceive and interpret the world (Wilson 2015), as well as CartoDB's positionality as a provider of the tools that enable such understandings, the failure of his maps to attend to or acknowledge the kind of connections mentioned above only works to reinforce problematic understandings of a range of social phenomena, especially the geographic dimensions of social media activity. It is this impoverishment of understanding that has led to a backlash against geotagged Twitter data as "possibly the worst metric of any modern scrapable dataset" (Field 2013; see also Goodspeed 2013). But, as the later analysis in this paper shows, viable alternatives exist to such a face-value approach to interpreting individual points on a map, which allow for a greater attention to the context of social media activity while also allowing for interpretations that support, rather than wholly dispense with, broader understandings of space and spatiality, and which allow for a substantively different understanding of questions of socio-spatial inequality.

III. Rethinking Social Media Mapping Relationally

While an often implicit adherence to Cartesian spatial ontology has continued to dominate the world of social media mapping, the last twenty to thirty years of geographic thought have seen a dramatic shift towards much more complex and situated understandings of space and spatiality that stands in direct contrast to Cartesian understandings of space. Drawing especially on Doreen Massey's (1991) early formulations around a 'global sense of place', the broad literature around what might be termed a '*relational* socio-spatial theory' conceives of space as networked, fragmented and processual, rather than as a kind of fixed container with defined boundaries and characteristics, such as single points or the more-or-less arbitrary Census-defined areal units typically used for spatial analysis. From reconceptualizations of globalization (Amin 2002) to a new focus on mobility as a fundamental, defining characteristic of contemporary life (Sheller and Urry 2006), a key tenet of this approach has been an inversion of Tobler's so-called 'first law of geography' – that all things are related, but near things are more related than far things. Instead, relational approaches suggest that "we cannot assume that local happenings or geographies are ontologically separable from those 'out there'" (Amin 2002: 386). By focusing on the social relations that recursively produce space and are in turn influenced by it, rather than simply privileging proximity in absolute, Cartesian space, Amin argues that we can begin to see "a subtle folding together of the distant and the proximate" (2007: 103).

The application of such relational insights to the similarly still-Cartesian world of GIS is, however, much easier said than done. Some of the earliest critiques of GIS remarked on the challenges of integrating multiple, competing representations of space into GIS, especially those that do not comply with "the logical rules used to relate geocoded information" (Sheppard 1995: 11). Similarly, Rundstrom (1995) argued that "At present, GIS does not capture relatedness, but constructs it. Relationships are reconstructed by assembling isolated pieces – in GIS terms, 'tuples,' 'data tables,' and 'layers' – of geographical information that have been torn from their context and 'corrected' separately" (47). Despite these longstanding critiques, Goodchild argues that little progress has been made on the front of integrating these alternative spatial ontologies and epistemologies into GIS, noting that "when GIS is adopted by indigenous peoples it is very much like the Cartesian GIS we know so well" (Goodchild 2006: 690). Even when more critical GIS work has departed from the GIS orthodoxy in order to argue for a more robust attention to questions of temporality and mobility, this work tends to couch such a shift as a de-emphasizing of space, rather than as a reconceptualization of space itself (cf. Kwan 2013).

So, much as has been the case for the last twenty years, the challenge is to mobilize fundamentally Cartesian data and forms of cartographic representation to understand the relational dimensions of social and spatial processes, moving from information about discrete 'sites' to understanding 'situations' and "the interrelationships between places" (Sheppard 1995: 11). As it relates to the matter of analyzing geotagged social media data, the question remains how to mobilize this data in such a way as to highlight its fundamentally relational character, rather than defaulting to a simplistic understanding that placing thousands or millions of dots on a map represents an analysis worth sharing. While each of these individual pieces of data remains fundamentally Cartesian in that they can be placed at a particular point on the earth's surface due to the attached geotag, it is through the other pieces of metadata attached to each point that allows for this relational perspective to be operationalized, in turn allowing more substantive and critically-oriented insights to be made.

A. Twitter as a Data Source

A more critical and relational approach to using geotagged social media data requires grappling with the data in a way that doesn't assume that the data, and in particular its explicit geographic reference, speaks for itself. It is important to not take the wealth of data contained within each individual data point – or, in this case, tweet – for granted by over-privileging the fact that each point can simply be placed on a map.

One of the key criticisms levied at the use of Twitter data in social research is its lack of representativeness. Given that only around 1 in 5 American adults, and 1 in 3 American teenagers are Twitter users (Pew Research Center 2015a, b) and something less than 5% of all tweets are geotagged, the data represent only a small sliver of the population, even in the United States. Furthermore, geotagged tweets are disproportionately skewed towards urban areas (Hecht and Stephens 2014), though some racial minorities in the US are actually over-represented relative to their proportion of the overall population (Pew Research Center 2015a). Nonetheless, Twitter, and geotagged tweets in particular, remain an incredibly limited data source in many respects, and because of these biases, is extremely problematic for purposes of predicting social behavior or inferring collective sentiment about a given issue (Lazer et al 2014; Ruths and Pfeffer 2014; Hargittai 2015). Because of these limitations in making inferences about the entire social world based on a very limited subset of individuals, Twitter data might better be seen as a real-time digital archive – with all of the attendant biases and limitations of more conventional archival sources – of individuals' everyday lives.

Focusing exclusively on geotagged data from Twitter, however, raises a number of other issues. While some methods exist for discerning some geographic references from tweets that aren't explicitly geotagged (Cheng et al 2010; Davis Jr. et al 2011; Mahmud et al 2012, 2014), focusing only on geotagged tweets ensures some level of certainty in the tweets actually having been created in the place to which they are tagged. And while the Twitter web interface allows for the tagging of tweets to places that the user may not be present in at that particular moment, this is a persistent issue across a range of other platforms (e.g., Flickr, Instagram, Wikipedia, Google Maps) whose user-generated data has been the focus of earlier social research. However, not all geotagged tweets are created equally, as there are varying degrees of accuracy or spatial resolution, from 'places' like points of interest, cities, counties, states and countries, to the much more precise latitude and longitude coordinates that are more often attached when tweeting from a mobile device, and which are largely unattainable through these alternative methods of location detection.

When tweets are tagged to areal units such as cities or states, the geographic coordinates are then interpreted as being the centroid of those areas. Because of this, only those data points with precise geographic coordinates are suitable for finer-grained analysis, such as at the urban scale. These point-based data, however, provide substantial advantages in urban analysis precisely because they aren't constrained by conventional areal units in the same way as census or other such data, which may not be available at finer scales. These individual points can then be put into relation with one another through a variety of methods, from aggregating to larger areal units of different kinds in order to find concentrations of tweeting, or by filtering larger datasets based on any of the pieces of metadata attached to each individual tweet beyond the geographic reference and time the tweet was created, from the user who created it and their life-history of previous tweets, shared themes in self-defined user descriptions or the number of tweets by each user and the number of other users they follow and are followed by.

It is these other pieces of metadata that form the backbone of a 'beyond the geotag' approach which simultaneously continues to make use of the explicitly geographic information attached to individual tweets, while also constructing a relational spatial understanding of this data that doesn't view each tweet as a kind of atomized individual divorced from its larger context. So building from Sheppard's broader argument that "[s]patiality can disrupt theories that have not taken it seriously"

(Sheppard, in Merriman et al 2012: 7), the paper now turns to demonstrating how the combination of this data with a more critical and relational socio-spatial perspective can yield alternative, more substantive insights than are possible when adhering to the overly simplistic spatial ontology of the geotag.

B. Rethinking the Geography of Ferguson-related Tweeting

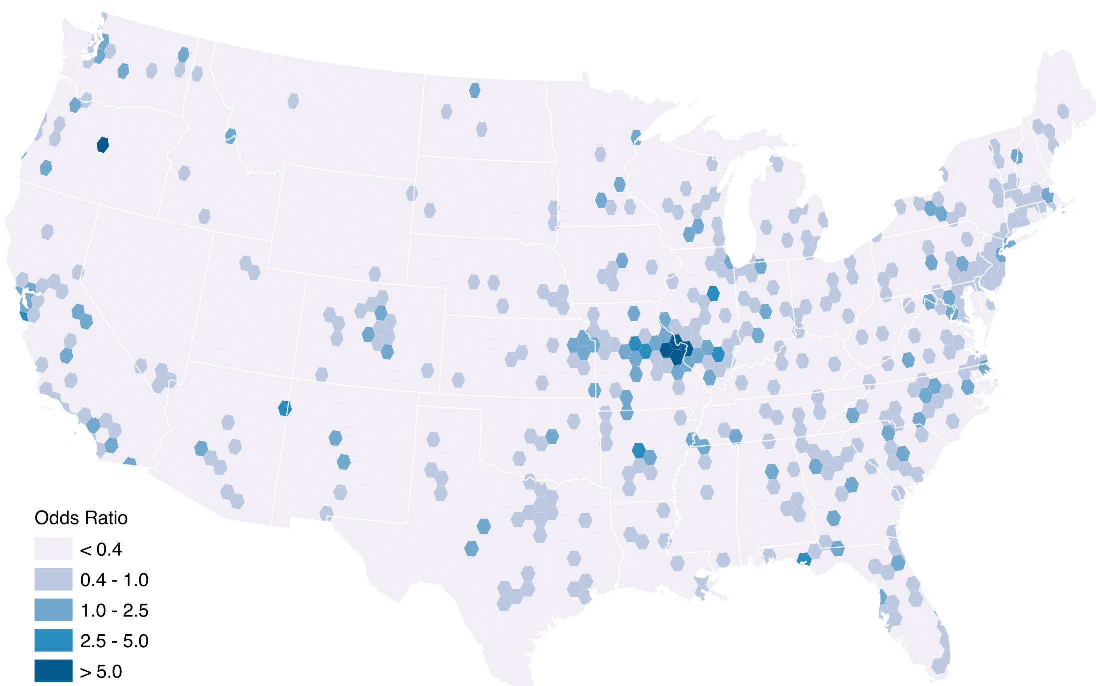
Although the spatial ontology of the geotag is an often-implicit conceptual premise, it tends to be expressed in very direct ways through the analysis and design of social media mapping projects. One of the most direct ways this occurs is through 'data dumps', which assume that simply plotting points on a map reveals some previously unforeseen truth, without much attention being paid to the quality or veracity of the data, or how it relates to other datasets. In this vein, Rogers' maps suffer from what Poorthuis and Zook (2015) call the problem of 'overplotting', in which countless points are simply layered on top of one another to the point that it's ultimately impossible to discern any meaningful spatial patterns from them. So while overplotting largely represents a flaw in visual design that can be adjusted without substantively changing one's overarching conception of socio-spatial relations, perhaps the most basic tenet of a relational approach to geotagged social media data is not to change the ways the data is represented, but to construct more complex and thoughtful ways of collecting and filtering data so as to more directly address the given questions at hand.

Arguably the easiest and most straightforward way to do this is by normalizing the dataset of, in this case, tweets by a baseline measure of Twitter activity. Even the most simplistic of normalization techniques avoids perpetuating the oft-cited problem of creating maps of Twitter that reproduce patterns of population density. But statistical measures like the odds-ratio, or location quotient as it tends to be known within spatial economics, allow for a more nuanced comparison of the phenomena in question – say, geotagged tweets about Ferguson within a given time frame – to the entire population – say, all geotagged tweets within the United States, regardless of topic, within the same time frame. Such approaches filter out much of the 'noise' associated with greater tweet density being highly correlated with greater population density, instead allowing for the researcher to highlight those locales that display unique concentrations of Twitter activity about a topic like Ferguson. The use of somewhat more complex statistical techniques, such as calculating the confidence interval of an odds-ratio, allows for a further filtering of noise by giving greater weight to those locations which simultaneous experience a greater relative amount of tweeting about the phenomena in question *and* a greater absolute amount of tweeting, preventing a counter-movement that gives too much analytical weight to greater proportional values in places with small total amounts of tweeting (see Poorthuis et al, forthcoming, for more discussion of these methods).

In order to demonstrate the utility of this approach, data was collected for a textual references to a variety of relevant terms in the seven days following August 9, 2014 at 3pm central time, the approximate time of Michael Brown's shooting in Ferguson. Tweets with references to "ferguson", "mikebrown", "handsup" and "dontshoot" were collected, totaling 56,110 tweets by 25,262 users, of which 51,145 tweets (created by 22,368 users) were within the United States. These tweets were then aggregated to uniformly-sized hexagonal cells and normalized by a 0.25% random sample of tweets (totaling 53,639 tweets) in the US during the same time period. By calculating the odds-ratio at the lower bound of the 95% confidence interval, which takes into account the levels of Ferguson-related tweets relative to the levels of tweeting one might otherwise expect to be in that place based on the random sample, we see a much different understanding of how the shooting of Michael Brown was reflected in the geography of social media.

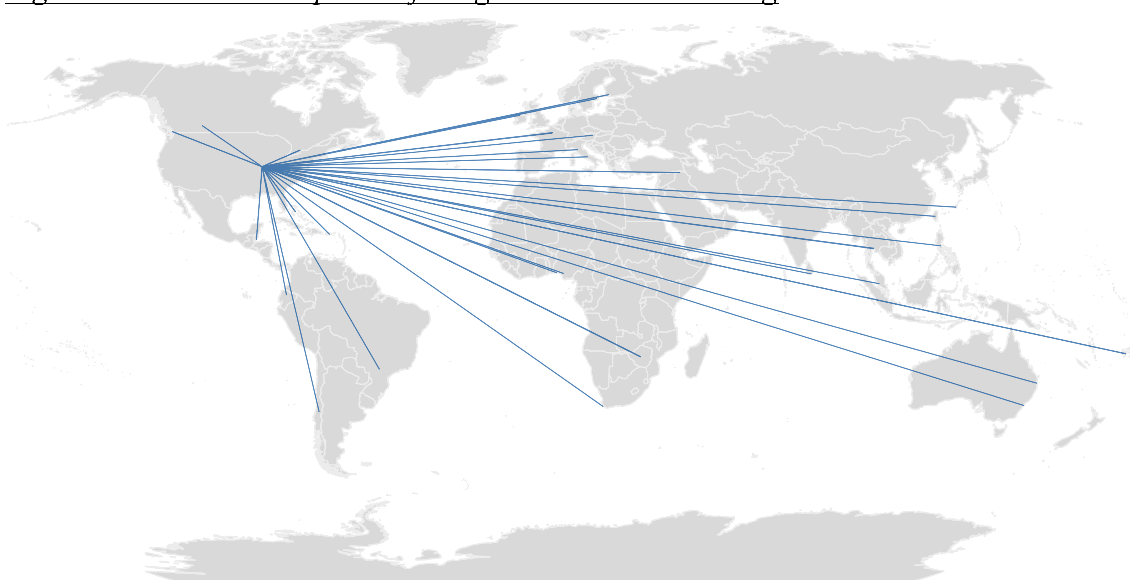
Indeed, rather than the globally dispersed flashes of Simon Rogers' animated map seen in Figure 2.1, Figure 2.2 shows that the epicenter of tweeting activity when accounting for baseline levels of tweeting is actually in the St. Louis metropolitan area. Indeed, the 4,606 geotagged tweets within the eleven county St. Louis metropolitan area are nearly as many tweets as the rest of the world outside the United States combined. When accounting for baseline levels of tweeting activity through normalization, this concentration of tweets in the St. Louis area is magnified by its typically lower amount of overall Twitter activity, while those highly populated and typically over-represented areas that show up prominently in Rogers' map – particularly the BosWash corridor in the northeast – are now much more muted. The four hexagonal areas that include most of the St. Louis metropolitan area each have confidence interval values greater than 5.5 – with the area including Ferguson and most of the city of St. Louis having a value greater than 17 – indicating that there was anywhere between five and seventeen times more tweets in the dataset of Ferguson-related tweeting from these areas than one would expect based on the random sample of tweets, values matched by only one other locale in the United States, which itself had much lower levels of absolute tweeting activity.

Figure 2.2: Normalized Map of Ferguson-related Tweets in the United States



So while Bonilla and Rosa were right in recognizing how the emergence of #Ferguson as a mediatized, virtual place is bound up in longer histories of racism and of anti-racist activism online and off, they err in arguing "social media users were able to show that '#Ferguson is everywhere'" (Bonilla and Rosa 2015: 12). Indeed, it is the somewhat subtle differentiation between the understanding of this spatial extensiveness as the death of distance and as a articulation of socio-spatial networks grounded in particular places at particular moments in time that characterizes the relational perspective (cf. Amin 2002). In order to demonstrate how this relational approach to data collection can be operationalized to go beyond the insights of Figure 2.2, the 4,965 Ferguson-related tweets from outside the United States were filtered according to the aforementioned other metadata fields to identify how many of the 2,894 users in that subset had also tweeted from within the United States at some point in time, signifying something of a more substantive connection with the events taking place in Ferguson. Of those users whose Ferguson-related tweets in the week following Michael Brown's murder were located outside of the United States, 1,002, or more than a third, had also created tweets within the United States in the three year period between July 2012 and June 2015, totaling 678,078 geotagged tweets.

Figure 2.3: Relational Spaces of Ferguson-related Tweeting



This data was then further filtered to identify those users who had previously tweeted from the area identified in Figure 2.2 as having the highest relative amount of Ferguson-related tweeting, which was expanded to include the census-defined St. Louis metropolitan area as a whole to simply data collection. While only 66 tweets from the original dataset of Ferguson-related tweets from outside the United States were created by users who have tweeted from the St. Louis metro area within a three-year period, these 33 users have created a total of 5,119 tweets from within this area. Figure 2.3 visualizes how the assumption of global interest in the events in Ferguson, as reflected in the aforementioned popular commentary on Simon Rogers' map, is somewhat more complex, with at least some of the globally dispersed tweeting also being related to – if not directly caused by – the very particular connections between individuals and places. It should also be mentioned that because of the potential for users to tag tweets to locations they aren't actually present in, initial efforts at data collection required filtering of users whose spatio-temporal tweeting patterns – namely tweet location changing drastically over the course of just a few hours – called into doubt whether they were actually located outside of the US at the time of their initial Ferguson-related tweets. While these users were scrapped from this exercise due to the questions they raised, if nothing else this only further points to the problems with declaring that globally-dispersed tweets are somehow a definitive marker of universal interest in a given issue such as Ferguson.

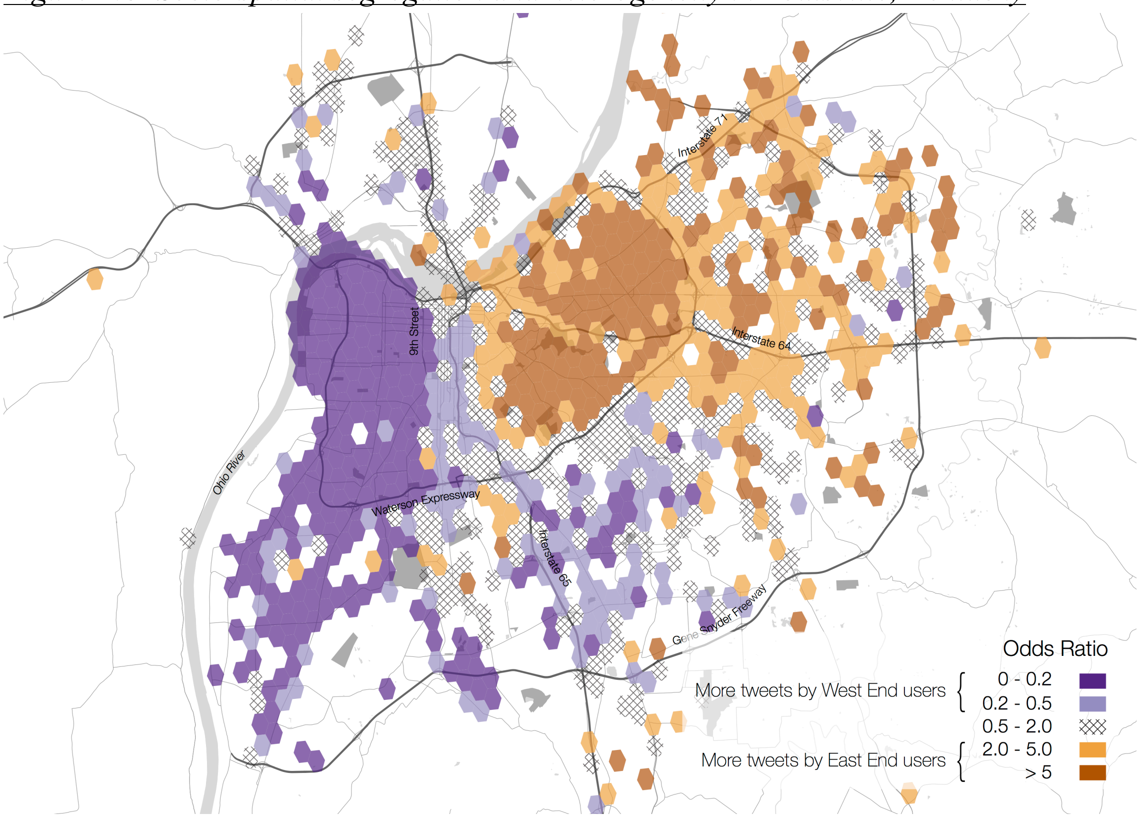
So rather than the implication being that the news of Michael Brown's shooting had transcended the confines of the place where the shooting occurred, we can instead understand both the fact that those in closer social and spatial proximity were more inclined to tweet about the events, as well as that the trans-local geographies of tweeting are not uniform across space or throughout society, but concentrated in other particular localities or nodes within the network, with some demonstrating stronger connections to the epicenter of these activities. These connections are only further evidenced by examining the qualitative data provided by the actual text of each tweet. For instance, one user in British Columbia tweeted "2,000 miles from home and we still get to watch the local news… #STL #Ferguson #ridiculous", while another in Fiji wrote "Reading the distressing #Ferguson tweets and wondering what I can do from here that will make a difference at home. Any ideas #expats?". While the spate of highly publicized instances of police or extrajudicial violence against unarmed African-Americans has occurred everywhere from St. Louis to Staten Island and suburban towns in Florida, these places aren't 'everywhere', but very particular nodes within an unevenly articulated, contingent and flexible network. Experiences of racism and police violence aren't universal, either socially or spatially, and neither are those moments of resistance to such injustices or the larger reactions they engender.

While the potentials for an analysis of geotagged tweets to uncover fundamental insights into the histories of racism or state violence in the United States or elsewhere are minimal, analyses like the one presented here do reinforce the notion that such processes are always fundamentally geographical and bound to particular understandings and enactments of place. And while Twitter, as well as other forms of social media and the internet more broadly, have allowed for a reconfiguration and rearticulation of social and spatial processes and relationships that often means an extension of the spatial reach of these processes, it hasn't eliminated these relationships and connections altogether. Indeed, extending the methods presented above, we can use this data to not only rethink the geographies of social media itself, but also to use a relational understanding of space to construct alternative understandings of socio-spatial inequalities and rethink the spatial categories that structure our ways of understanding and intervening in these fundamentally unequal landscapes.

C. Mapping Socio-spatial Inequalities Using Geotagged Social Media Data

Like St. Louis, Louisville, Kentucky is a city characterized by stark racial inequalities embedded in the urban landscape and the socio-spatial imaginary of its residents. In St. Louis, the so-called 'Delmar Divide' separates predominantly black North St. Louis from its predominantly white counterparts to the south, while in Louisville the '9th Street Divide' separates the city's largely poor and black West End from its more affluent and white counterparts to the east (cf. Crutcher 2013; Harlan 2014). In these cities, like many others, socio-spatial inequalities can often be seen through unevenness in the densities of data shadows in these different areas. For example, in Louisville the West End has only about 40% as many tweets as a comparable area of the city's East End. But as was demonstrated in the previous section through the analysis of Ferguson-related tweets outside of the United States by those with connections to the St. Louis area, a relational perspective attunes us to the fact that the connection between a single data point in a particular place belies a range of other social and spatial processes. As the more thorough analysis presented in Shelton et al (forthcoming) explicates, the seemingly fundamental separation between the West End and East End that pervades locals' understanding of the city's landscape can be rethought by analyzing the connections between people and places and their movements through the city over time as seen through geotagged tweets.

Figure 2.4: Socio-spatial segregation and heterogeneity in Louisville, Kentucky



Source: Shelton et al (forthcoming)

Starting from approximately two years of geotagged tweets in Louisville, data was filtered to create two parallel databases representing tweets from those users who could be defined as 'belonging' to either the West End or the East End[5]. After identifying these two separate user groups, we can extend the methods of normalization discussed in the previous section by comparing these two groups against each other using an odds-ratio, rather than to a baseline measure of the Twitter 'population'. Figure 2.4 above visualizes a comparison between West End and East End Twitter users in Louisville through an adjusted count of tweets meant to account for the existence of so-called 'power users' who might exert a disproportionate influence on such a dataset. As such, the map represents those places that are more socially segregated based on the mobility of the roughly 1,400 individuals in our dataset, with places dominated by West End users having values approaching 0 and represented with purple hexagons, and places with values increasing beyond 1 showing a dominance of East End users and represented by orange hexagons. In the odds-ratio measure, a value of exactly 1 represents parity between the two user groups, in this case signifying some level of social heterogeneity or co-presence in space between the two user groups. A cushion on either side of this value is represented in Figure 4 with the cross-hatched pattern to show those places which have relative heterogeneity compared to the rest of the city.

[5] In order to be included in these datasets for either user group, users were required to have created at least 40 tweets within the defined neighborhood boundary, and have those tweets represent more than 50% of their total geotagged tweeting activity within the city. This ensures that users 'belong' to only one of the two areas, and that they have some sustained presence in that neighborhood as represented through a significant density of content.

While those areas used to define the West End and the East End for the purposes of data collection show a general dominance of users from those areas, the map is most significant in that it offers a counterpoint to the dominant popular socio-spatial imagination of the West End as fundamentally separate and apart from the rest of the city, as represented in the colloquial invocation of the '9th Street Divide'. Though 9th Street is visible as roughly the point at which the dominant concentration of West End tweeting gives way to greater heterogeneity in the downtown area, the smattering of both heterogeneous areas and scattered areas of West End tweeting throughout the rest of the city serve as a counterpoint to the simplistic understanding of the West End as 'walled off' from the rest of the city. Instead, one can see based on the minimal areas of East End dominant tweeting throughout the rest of the city that it's actually those East End users who are more spatially segregated within the city. And while, like the earlier analysis which began only with the isolated, individual data points, this analysis begins from the problematic understandings of this socio-spatial divide, it is through such a relational analysis that these spatial ontologies can be challenged and rethought.

That is, Figure 4 shows that rather than being a kind of informational 'blank spot on the map' divorced from the rest of the city, the West End is actually thoroughly represented in Louisville's digital footprint, just not within the conventional boundaries of the West End itself. Though the West End may have 40% less tweeting activity within its border than the East End, we can see that a significant portion of those tweets within the East End are actually being created by users who we can classify as belonging to the West End. In other words, the fundamental inequalities of needing to travel significant distances across the city for work, school or consumptive activities leads to the illusion that the West End is largely cut off from these new forms of digital representation. Instead, we can see that the West End is fundamentally implicated in the production of such informational content through the movements of its residents into other parts of the city where such content ends up being produced in greater concentrations. So even though we were able to identify just 662 users belonging to the West End as compared to the East End, this slightly smaller number of users actually created a substantially larger number of geotagged tweets (n=398,432) as compared to their counterparts (n=274,338).

As is further explicated in Shelton et al (forthcoming), comparing these spatial footprints allows us to reconstruct empirically grounded understandings of the West End as spatially extensive and fundamentally interconnected with the rest of the city. But like the analysis of Ferguson-related tweeting presented above, this data can still only provide a very partial understanding of *why* people move through the city the way they do, though it can still be extremely useful for understanding how social and spatial inequalities manifest themselves in practice. Perhaps more importantly, combining this data with a sensitivity to the complexities of these places and the way we imagine them allows for a counterpoint to more dominant discourses around the 9th Street Divide – not to mention similar divides in other American cities – that serve to isolate and pathologize the West End and its residents, reinforcing the harmful policies of social and spatial segregation that are generative of such imaginaries in the first place.

Even though filtering down to the local level in both the cases of Ferguson and Louisville can mean dealing with a fairly small number of users, it is this process of identifying the connections between people and places that are embedded within these larger datasets, but which are largely ignored by the 'data dumps' created by Simon Rogers and others, that represents significant promise for extending geographical research utilizing social media data in a way that accords with broader conceptual trends within the discipline. So far from reinscribing simplistic understandings of the globalizing and geography-nullifying effects of new information and communication technologies like Twitter, these new sources of data allow for a demonstration of the deepening – if also complexification – of the connections between the offline, material world and its digital reflections (Graham 1998).

IV. Conclusion

Geographically-referenced data drawn from social media platforms is surely limited in its usefulness for understanding many social and spatial processes. But, as this paper has argued, these limitations do not represent a *fundamental* challenge to the use of this data for many purposes in social – and more specifically, geographical – research. Instead, those popular attempts at mapping social media data to provide insight into broader social processes are often limited not by inherent flaws in the data, but by an overly simplistic spatial ontology and a lack of attention to the context the data is embedded in and the process by which we can make meaning out of it.

By bringing together work on relational socio-spatial theory and critical approaches to GIS to apply to these new datasets, substantively different understandings of social and spatial processes can be developed, as was highlighted through the cases of tweeting about the events in Ferguson, Missouri in August 2014 and of urban socio-spatial inequalities in Louisville, Kentucky. Rather than simply plotting points on a map in order to either marvel at the interest in these events or point to the obvious unevenness in these digital data shadows, we can apply a variety of statistical, cartographic and data collection methods in order provide a more nuanced understanding of both the geography of reactions to Michael Brown's killing in Ferguson and the everyday movements of people from different parts of the city by linking each otherwise-isolated data point to one another. Like other nascent attempts – explicit or otherwise – to use these new sources of digital data to rethink the boundedness of conventional Cartesian spatial categories (cf. Cranshaw et al 2012; Stefanidis et al 2013; Shelton et al 2014; Shelton et al, forthcoming), this paper has tried to demonstrate how these point-based datasets can actually be incorporated into understandings of socio-spatial processes that are fundamentally about the connectedness of people and places, even those that might be distant in absolute space, highlighting the fundamentally place-based nature of social processes, while understanding these places and processes to be fluid and spatially extensive.

While this research has pointed towards the utility of different methods for approaching this data relationally, future work should look to a couple of key issues in order to push forward the development of a relational approach to social media mapping and analysis. First, processes for filtering data into smaller subsets in order to draw out the particular place-based social networks around different issues should be expanded to focus on multiple sites and scales, from the global to the neighborhood level. While the Ferguson example presented here draws attention to how a small proportion of non-local tweeting about the events is still connected to the St. Louis metropolitan area, it is possible that multiple such connections between places and over time might exist within this same framing, perhaps connecting Ferguson to more recent incidents of racist violence in the United States. Second, future work should continue to more directly address substantive questions of inequality, as the Louisville case study points towards. While such popular and controversial discussions provide fodder for testing new concepts and methods, the potentials of this data to say something meaningful about people's everyday lives remains relatively unexplored, despite the persistent hype around this data as providing novel insight into such matters.

V. References

Amin, Ash. 2002. "Spatialities of Globalisation." *Environment and Planning A* 34 (3): 385–399.

Amin, Ash. 2007. "Re-thinking the Urban Social." *City* 11 (1): 100–114.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine* 15(7).

Barnes, Trevor J. 2013. "Big Data, Little History." *Dialogues in Human Geography* 3 (3): 297–302.

Bonilla, Yarimar, and Jonathan Rosa. 2015. "#Ferguson: Digital Protest, Hashtag Ethnography, and the Racial Politics of Social Media in the United States." *American Ethnologist* 42 (1): 4–17.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–679.

Brownlee, John. 2014. "Watch How the Twitter Conversation of #Ferguson Spread Across America". *FastCo Design*. 18 August. Available from: http://www.fastcodesign.com/3034472/infographic-of-the-day/watch-how-the-twitter-conversation-of-ferguson-spread-across-america

Capps, Kriston. 2014. "Twitter Made an Amazing Map of Twitter Going Nuts Over Ferguson". *CityLab.* 14 August. Available from: http://www.citylab.com/crime/2014/08/twitter-made-an-amazing-map-of-twitter-going-nuts-over-ferguson/376119/

CartoDB. 2014. "CartoDB Twitter Maps in Today, Yahoo, ESPN, Time, CNBC, USA Today, Daily News, Huffington Post…". *CartoDB Blog.* 28 June. Available from: http://blog.cartodb.com/cartodb-twitter-maps-in-today-yahoo-espn-time-cnbc/

Chen, Xiang, and Xining Yang. 2014. "Does Food Environment Influence Food Choices? A Geographical Analysis through 'tweets.'" *Applied Geography* 51: 82–89.

Cheng, Zhiyuan, James Caverlee, and Kyumin Lee. 2010. "You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users." In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 759–768.

Crampton, Jeremy W. 2010. *Mapping: A Critical Introduction to Cartography and GIS*. John Wiley & Sons.

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb." *Cartography and Geographic Information Science* 40 (2): 130–139.

Cranshaw, Justin, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. "The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City." In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, June, pp. 58–65.

Crooks, Andrew, Arie Croitoru, Anthony Stefanidis, and Jacek Radzikowski. 2013. "#Earthquake: Twitter as a Distributed Sensor System." *Transactions in GIS* 17 (1): 124–147.

Crutcher, Dan. 2013. “A Tale of Two Cities”. *Louisville Magazine* (March): 25-29.

Crutcher, Michael, and Matthew Zook. 2009. “Placemarks and Waterlines: Racialized Cyberscapes in Post-Katrina Google Earth.” *Geoforum* 40 (4): 523–534.

Curry, Michael R. 1995. “On Space and Spatial Practice in Contemporary Geography.” In *Concepts in Human Geography*, edited by Carville Earle, Kent Mathewson, and Martin S. Kenzer. Rowman and Littlefield. pp. 3–32.

Davis, Jr., Clodoveu A., Gisele L. Pappa, Diogo Rennó Rocha de Oliveira, and Filipe de L. Arcanjo. 2011. “Inferring the Location of Twitter Messages Based on User Relationships.” *Transactions in GIS* 15 (6): 735–751.

Field, Kenneth. 2013. “The flawless map”. *Cartonerd*. 17 December. Available from: http://cartonerd.blogspot.com/2013/12/the-flawless-map.html

Field, Kenneth. 2014. “I'm wondering when people will realise the animated ectoplasm twitter maps don't actually show anything http://t.co/SJVYLyBn1F” [Tweet]. 17 June. Available from: https://twitter.com/kennethfield/status/478775510386741248.

Fischer, Eric. 2010. “Locals vs. Tourists”. Available from: https://www.flickr.com/photos/walkingsf/sets/72157624209158632/

Fischer, Eric. 2014. “Making the most detailed tweet map ever”. *Mapbox Blog*. 3 December. Available from: https://www.mapbox.com/blog/twitter-map-every-tweet/

Fung, Brian. 2014. “Watch Twitter explode along with Ferguson”. *The Washington Post.* 14 August. Available from: http://www.washingtonpost.com/blogs/the-switch/wp/2014/08/14/watch-twitter-explode-along-with-ferguson

Goodspeed, Robert. 2013. “The Limited Usefulness of Social Media and Digital Trace Data for Urban Social Research.” In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*. pp. 1-4.

Goodchild, Michael. 2006. “GIScience Ten Years After Ground Truth.” *Transactions in GIS* 10 (5): 687–692.

Graham, Mark, and Taylor Shelton. 2013. “Geography and the Future of Big Data, Big Data and the Future of Geography.” *Dialogues in Human Geography* 3 (3): 255–261.

Graham, Mark, and Matthew Zook. 2011. “Visualizing Global Cyberscapes: Mapping User-Generated Placemarks.” *Journal of Urban Technology* 18 (1): 115–132.

Graham, Mark, and Matthew Zook. 2013. “Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web.” *Environment and Planning A* 45 (1): 77–99.

Graham, Stephen. 1998. "The End of Geography or the Explosion of Place? Conceptualizing Space, Place and Information Technology." *Progress in Human Geography* 22 (2): 165–185.

Hargittai, Eszter. 2015. "Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites." *The Annals of the American Academy of Political and Social Science* 659 (1): 63–76.

Harlan, Chico. 2014. "In St. Louis, Delmar Boulevard is the line that divides a city by race and perspective". *The Washington Post*. 22 August. Available from: http://www.washingtonpost.com/national/in-st-louis-delmar-boulevard-is-the-line-that-divides-a-city-by-race-and-perspective/2014/08/22/de692962-a2ba-4f53-8bc3-54f88f848fdb_story.html

Hawelka, Bartosz, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. "Geo-Located Twitter as Proxy for Global Mobility Patterns." *Cartography and Geographic Information Science* 41 (3): 260–271.

Hecht, Brent, and Monica Stephens. 2014. "A Tale of Cities: Urban Biases in Volunteered Geographic Information." In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. pp. 197–205.

Hill, Andrew. 2015. "In defense of burger cartography: Or, time to fall in love with maps all over again". 28 March. Available from: http://andrewxhill.com/blog/2015/03/28/in-defense-of-burger-cartography/

Kirsch, Scott. 1995. "The Incredible Shrinking World? Technology and the Production of Space." *Environment and Planning D: Society and Space* 13 (5): 529–555.

Kitchin, Rob. 2014a. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1).

Kitchin, Rob. 2014b. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.

Kwan, Mei-Po. 2013. "Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility." *Annals of the Association of American Geographers* 103 (5): 1078–1086.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176): 1203–1205.

Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook. 2013. "Mapping the Global Twitter Heartbeat: The Geography of Twitter." *First Monday* 18 (5).

Llorente, Alejandro, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. 2014. "Social Media Fingerprints of Unemployment." *arXiv.* Available from: http://arxiv.org/abs/1411.3140.

Lotan, Gilad. 2011. "Data Reveals That 'Occupying' Twitter Trending Topics is Harder Than it Looks!". 12 October. Available from: http://giladlotan.com/2011/10/data-reveals-that-occupying-twitter-trending-topics-is-harder-than-it-looks/

Mahmud, Jalal, Jeffrey Nichols, and Clemens Drews. 2012. "Where Is This Tweet From? Inferring Home Locations of Twitter Users." In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. pp. 511–514.

Mahmud, Jalal, Jeffrey Nichols, and Clemens Drews. 2014. "Home Location Identification of Twitter Users." *ACM Transactions on Intelligent Systems and Technology* 5 (3): 1–21.

Massey, Doreen. 1991. "A Global Sense of Place." *Marxism Today* 35 (6): 24–29.

Massey, Doreen. 1999. "Philosophy and Politics of Spatiality: Some Considerations. The Hettner-Lecture in Human Geography." *Geographische Zeitschrift* 87 (1): 1–12.

Merriman, Peter, Martin Jones, Gunnar Olsson, Eric Sheppard, Nigel Thrift, and Yi-Fu Tuan. 2012. "Space and Spatiality in Theory." *Dialogues in Human Geography* 2 (1): 3–22.

Mitchell, Lewis, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. 2013. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place." *PLoS ONE* 8 (5).

NBC News. 2014. "As the debate and fires raged, Twitter's "heat map" shows an engaged American public. http://nbcnews.to/1vJVqIa" [Tweet]. 25 November. Available from: https://twitter.com/NBCNews/status/537270894251749377

Pew Research Center. 2015a. "Social Media Update 2014". Available from: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf

Pew Research Center. 2015b. "Teens, Social Media and Technology Overview 2015". Available from: http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf

Poorthuis, Ate, and Matthew Zook. 2015. "Small Stories in Big Data: Gaining Insights From Large Spatial Point Pattern Datasets." *Cityscape* 17 (1): 151–160.

Poorthuis, Ate, Matthew Zook, Taylor Shelton, Mark Graham, and Monica Stephens. Forthcoming. "Using Geotagged Digital Social Data in Geographic Research."

In *Key Methods in Geography (3rd Edition)*, edited by Nick Clifford, Shaun French, Meghan Cope, and Thomas Gillespie. Sage.

Rundstrom, Robert A. 1995. “GIS, Indigenous Peoples, and Epistemological Diversity.” *Cartography and Geographic Information Systems* 22 (1): 45–57.

Ruths, Derek, and Jürgen Pfeffer. 2014. “Social Media for Large Studies of Behavior.” *Science* 346 (6213): 1063–64.

Sheller, Mimi, and John Urry. 2006. “The New Mobilities Paradigm.” *Environment and Planning A* 38 (2): 207–226.

Shelton, Taylor, Ate Poorthuis, Mark Graham, and Matthew Zook. 2014. “Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of ‘Big Data.’” *Geoforum* 52: 167–179.

Shelton, Taylor, Ate Poorthuis, and Matthew Zook. Forthcoming. “Social Media and the City: Rethinking Urban Socio-Spatial Inequality Using User-Generated Geographic Information.” *Landscape and Urban Planning*.

Shelton, Taylor, Matthew Zook, and Mark Graham. 2012. “The Technology of Religion: Mapping Religious Cyberscapes.” *The Professional Geographer* 64 (4): 602–617.

Sheppard, Eric. 1995. “GIS and Society: Towards a Research Agenda.” *Cartography and Geographic Information Science* 22 (1): 5–16.

Sheppard, Eric. 2001. “Quantitative Geography: Representations, Practices, and Possibilities.” *Environment and Planning D: Society and Space* 19 (5): 535–554.

Soja, Edward W. 1980. “The Socio-Spatial Dialectic.” *Annals of the Association of American Geographers* 70 (2): 207–225.

Stefanidis, Anthony, Amy Cotnoir, Arie Croitoru, Andrew Crooks, Matthew Rice, and Jacek Radzikowski. 2013. “Demarcating New Boundaries: Mapping Virtual Polycentric Communities through Social Media Content.” *Cartography and Geographic Information Science* 40 (2): 116–129.

Sui, Daniel. 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.

Sullivan, Gail. 2014. “How Facebook and Twitter control what you see about Ferguson”. *The Washington Post*. 19 August. Available from: http://www.washingtonpost.com/news/morning-mix/wp/2014/08/19/how-facebook-and-twitter-control-what-you-see-about-ferguson/

Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. 2012. “Geography of Twitter Networks.” *Social Networks* 34 (1): 73–81.

Warf, Barney, and Santa Arias. 2008. *The Spatial Turn: Interdisciplinary Perspectives*. Taylor & Francis.

Widener, Michael J., and Wenwen Li. 2014. "Using Geolocated Twitter Data to Monitor the Prevalence of Healthy and Unhealthy Food References across the US." *Applied Geography* 54: 189–197.

Wilson, Matthew W. 2015. "Morgan Freeman Is Dead and Other Big Data Stories." *Cultural Geographies* 22 (2): 345–349.

Wyly, Elvin. 2009. "Strategic Positivism." *The Professional Geographer* 61 (3): 310–322.

Wyly, Elvin. 2011. "Positively Radical." *International Journal of Urban and Regional Research* 35 (5): 889–912.

Wyly, Elvin. 2014. "The New Quantitative Revolution." *Dialogues in Human Geography* 4 (1): 26–38.

Chapter 3
Mapping the Data Shadows of Hurricane Sandy: Uncovering the Socio-spatial Dimensions of ‘Big Data’

Published as:
Shelton, Taylor, Ate Poorthuis, Mark Graham, and Matthew Zook. 2014. “Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of ‘Big Data.’” *Geoforum* 52: 167–179.

I. Introduction

Digital social data are now practically ubiquitous. This data is nowhere more visible than on the Internet, as over two and half billion people currently both actively produce content, and leave behind all manner of transactional records, from comments and 'likes' on Facebook to the different products one has viewed and purchased on Amazon. In addition to online traces, people, buildings, roads, machines, plants and animals, alike, are increasingly augmented with sensors and software algorithms that produce electronic records of all manner of social, economic, political and environmental processes. These sources of digital data combine to create what we call *'data shadows'* (Zook et al 2013; Graham 2013; Graham and Shelton 2013), or the imperfect representations of the world derived from the digital mediation of everyday life. As these datasets grow exponentially, researchers, politicians, and the private sector have begun to focus on how 'big data' might allow potentially unprecedented insights into our world (Hey and Trefethen 2003; Anderson 2008; Floridi 2012).

Much of the 'big data' being produced online through social media has a significant amount of geographic information attached to it, often in the form of latitude and longitude coordinates known as 'geotags', which provide the means for new ways of doing, creating, making, and enacting geography. This process of attaching geographic coordinates to user derived digital content – often referred to as the geoweb – means that big data shadows are intimately connected to the material lived geographies from which they were produced. As such, social media has evolved beyond a simple online repository of conversations, networked interactions, and sites for the consumption of media, and is instead a dynamic record of when and how we move through and act in space, linked to other individuals and actions co-existing with us in those spaces. It is this connection between the geographies of online big data and the material processes they represent, and in turn impact, that we interrogate in this paper. In other words, what can big data from geographically referenced social media reveal about material processes and practices? And what can our pre-existing knowledge about such material processes and practices tell us about the underlying spatialities of big data?

In order to call attention to the interrelations between the material world and its connections to the virtual practices of what might more accurately be called '*geo*social media', we highlight a case study of Twitter activity in the wake of Hurricane Sandy, which struck the eastern seaboard of the United States in late October 2012. The second-most costly storm in US history behind only Hurricane Katrina, Sandy wreaked havoc on New York City's infrastructural systems, creating iconic images of flooded subway tunnels and roadways, dangling construction cranes and a blacked-out Lower Manhattan. In spite of these disruptions, the material effects of Sandy on New York City and the lives of people living in affected areas were clearly reflected in their online social media activities, as well as in the online activities of people living thousands of miles away. As such, the hurricane offers an accessible way to describe the variety of socio-spatial relationships embodied in these big data shadows.

This paper argues that Hurricane Sandy offers a useful lens for understanding the digital data shadows produced by intensely material phenomena. Applications of big geosocial media data are increasing common throughout a range of activities beyond just disaster response, from urban planning to market research to political activism, and this case study provides the basis for a series of broad methodological and theoretical interventions into research on big data and user-generated geographic information. Methodologically speaking, rather than simply focusing on how massive databases are causing necessary and irreversible shifts in social practices or producing unprecedented insights into the world around us, we instead argue that it is more productive to analyze how small subsets of big data, especially georeferenced social media information, can reveal a broader range of social, economic, political, and even environmental geographies. Utilizing a mix of qualitative and quantitative methods, we uncover both broad spatial patterns within this data, as well as understand how these data reflect the lived experiences of the people who are creating it. Conceptually, we seek to fill a gap in previous studies of the geoweb, which have often avoided explicitly theorizing the nature of socio-spatial relations. Building on Jessop et al's (2008) Territory-Place-Scale-Network (or TPSN) framework for understanding socio-spatial relationships, we analyze the territorial, platial, scalar and networked dimensions of digital data shadows to highlight the polymorphous and complex spatialities of user-generated content. This allows for a greater consideration of the relational geographies of big data and geosocial media, which have largely been neglected in the literature to this point, while retaining an attention to more conventional ways of understanding the spatialities of this data.

In the following sections, we first review the relevant literature, focusing on conceptualizations and problematizations of big data. We then turn to understanding how big and user-generated data sources have been utilized in disaster response situations, before discussing the potential for new theorizations of socio-spatial relations in studies of the geoweb. This is followed by a discussion of our data collection and methods, with attention to the potentials of using geotagged tweets for social and spatial analysis. In the penultimate section, we turn to the case of Hurricane Sandy and use a series of cartographic visualizations to highlight the variegated and polymorphous nature of socio-spatial relations represented by Sandy's data shadows. Finally, we discuss the possibilities for and limitations of future studies of big data shadows.

II. Contextualizing 'Big Data' and Geosocial Media

This work is framed within the context of an important shift occurring in the social sciences: the emergence of 'big data', or what has been referred to as the 'fourth paradigm' of scientific research (Hey et al 2009; Mayer-Schonberger and Cukier 2013). Big data's proliferation throughout the popular press as a buzzword comes with many different definitions, and it is important to recognize that it refers not just to a quantitative increase in the size of the datasets being analyzed, but also qualitative shifts in the ways we approach the study of society (boyd and Crawford 2012). These shifts include an increase in the scope of the data being collected, the speed and timeframe at which it is collected, and the notion that otherwise unrelated datasets might be cross-referenced and analyzed to produce some meaningful insight (Kitchin 2013).

Perhaps the most prominent proponent of this new data-driven science has been Chris Anderson, the former editor of *Wired Magazine*, who sees the proliferation and availability of these new datasets as a way to generate more insightful, useful, accurate, or true results than more conventional specialists or domain experts who carefully develop hypotheses and research strategies in order to understand a given phenomena – heralding 'the end of theory' (Anderson 2008). Anderson's notion has entered not only the popular imagination, but also the research practices of corporations, states, journalists and academics (Lazer et al 2009; Leetaru 2011; Issenberg 2012; Lohr 2012; see also Torrens 2010 for a geographic perspective), driven by the idea that the data shadows of people, machines, commodities, and even nature, can reveal difficult-to-understand social processes, simply by applying sufficient computing power to these massive amounts of data. In other words, researchers no longer need to speculate and hypothesise; they simply need to possess enough data and allow algorithms to lead them to important patterns and trends in social, economic, political, and environmental relationships.

This kind of naïve technological determinism echoes a similar argument made a decade earlier about the so-called 'death of distance' (Cairncross 1997) brought by the internet, which itself stimulated a range of more nuanced theoretical and empirical works on the geography of the internet. Anderson's hyperbole around the end of theory has also given rise to a range of critical responses from social scientists of all types. This critical approach to big data has been especially pronounced amongst those scholars studying the geographic contours of user-generated internet content, as notions of big data frequently incorporate elements of what have variously been called the geoweb or volunteered geographic information (Goodchild 2007; Elwood 2008; Elwood et al 2012). Nonetheless, scholars are just now beginning to employ social media data to ask substantive questions about the geographies of production, use and consumption of big data (Takhteyev et al 2012; Graham et al 2013; Tsou and Leitner 2013).

Two primary criticisms of such big data analyses have been their failure to attend to persistent methodological issues and their overblown claims to be able to deduce significant meaning out of data without relying on pre-existing theoretical frameworks. In arguably the most visible critique of big data so far, danah boyd and Kate Crawford note that “Big Data and whole data are also not the same” (boyd and Crawford 2012: 669). Similarly, Muki Haklay (2012) has warned that too often, analysis of big social media datasets tends to privilege the perspectives of so-called ‘outliers’, rather than incorporating a representative sample of the population. So while big data can capture a whole host of social processes that were previously difficult to study because of their transactional nature[6], it remains partial and biased in important

[6] Transactional data is used to refer to data describing events, which until recently were not readily accessible. This could quite literally include data describing a financial transaction or purchase at a store, or more loosely the kind of social media data we discuss in this paper. Of course, for this kind of data to become useful when cross-referenced with other databases, these transactions must be digital and

ways that should qualify any findings from such research (Manovich 2011; Ruppert et al 2013).

Meanwhile, proponents of big data have also been critiqued for their relatively naïve claims to have transcended the need for any domain expertise in the subjects they analyze (Graham 2012). For example, the physicist Geoffrey West has supposedly 'solved the city' using mathematical approaches, without having ever read any work in urban studies (Lehrer 2010), while others have used similar databases of Twitter activity to revive the 'death of distance' thesis (Leetaru et al 2013). It should be noted, however, that others, especially geographers, have been a good bit more cautious. For instance, Miller (2010) argues that data-driven approaches have much to benefit from the inclusion of more conventional domain expertise, while more traditional approaches to social science can benefit from the improved hypothesis generating capabilities of data mining.

A. Social Media and Crowdsourcing Disaster Response

One realm in which the role of big and user-generated data has generated massive amounts of attention has been disaster response (cf. Goodchild and Glennon 2010; Li and Goodchild 2010; Liu and Palen 2010; Roche et al 2013). While key players in this space, such as Ushahidi and the crisis mapping community, developed in a more-or-less ad-hoc manner in order to respond to disasters such as the 2010 Haitian earthquake,

---

automatically registered, which, for instance, would tend to exclude individuals whose economic activities are predominantly informal or cash-based. It would similarly exclude anyone who chooses not to participate in social media or other similar services.

more established institutions, including states and international NGOs, are similarly promoting the potentials of these new data sources and their analysis for responding to disaster or crisis situations. For example, The Red Cross has actively been collecting tweets about disaster situations (Red Cross n.d.), while the United States Geological Survey has been using its 'Did You Feel It?' online reporting tool to crowdsource reports about the intensity of earthquakes for over a decade (Wald et al 1999; Wald and Dewey 2005). Other less systematic examples include the much-publicized efforts of Newark, New Jersey mayor Cory Booker to personally respond to unfulfilled service requests during a major blizzard in that city, following attempts by residents to use Twitter to encourage a response (Gregory 2010).

While there are many such examples of success in harnessing this kind of data for disaster response, two important contributions to this discussion from a geographic perspective raise questions about their efficacy. Crutcher and Zook (2009) and Zook et al (2010), discussing the use of social media in response to Hurricane Katrina and the Haitian earthquake respectively, argue that patterns of adoption and utilization of such technologies in disaster response have largely followed long-standing patterns of socio-spatial inequality, producing uneven data shadows that don't reflect the on-the-ground realities following disasters. This is driven, or at least exacerbated, by the fact that such disasters typically represent the failures or inadequacies of state-based disaster relief, leading to a greater number of citizens taking an active role in the production of information about such events (cf. Leszczynski 2012, on the relationship between the rollback of state functions and the production of geographic information).

In the case of Hurricane Katrina, Crutcher and Zook were able to show that the production of user-generated, geotagged reports tended to be associated with wealthier, whiter, more tourist-oriented locations within New Orleans, despite the greatest effects of the storm being felt in predominantly poor and black areas, such as the Lower Ninth Ward. As the later case of the Haitian earthquake of 2010 demonstrates, however, such disasters *can* serve to stimulate greater attention to the production of user-generated geographic information in and about such marginalized places. Indeed, some of the most striking examples of the volunteer effort following the earthquake are those that demonstrate the lack of codified and widely-accessible geographic information about the country prior to the earthquake, and the explosion of information produced following it in order to aid in the recovery effort (cf. Zook et al 2010 for visualizations of the growth in user-generated Google placemarks following the earthquake, or ITO World 2010 for a time-lapse video of edits to OpenStreetMap). Nonetheless, these findings demonstrate that while such participatory, citizen-driven and technology-centric efforts have great potential to aid in disaster situations, these solutions are only ever partial, both in terms of participation and assistance, and are no replacement for more coordinated 'on the ground' relief efforts.

As such, the case study of Hurricane Sandy used in this paper represents an important opportunity to revisit these earlier findings in a different context. Apart from the contextual differences between New York City and New Orleans or Port-au-Prince, one major difference between Sandy and the earlier cases hinges on technology. While Google Earth had just been released when Hurricane Katrina struck the Gulf Coast in 2005, and the Haitian earthquake represented something of a test case for technology-based disaster response at a distance, the nearly 20 million tweets about Hurricane Sandy (Twitter n.d.) provide a sufficiently robust source of data to map the data shadows of the storm. This wealth of user-generated data can help us in better understanding the connections between the material world and its virtual representations. It also allows us to articulate a more coherent conceptualization of the spatialities of these data shadows in order to counter the dominant popular discourse that sees big data as an objective and normatively superior way of understanding the world, and to fill conceptual gaps that remain in the critical literature on these issues.

But as Kate Crawford (2013) points out, referencing both the case of Hurricane Sandy and the aforementioned paper by Crutcher and Zook, one cannot rely solely on social media content to reveal where the most damage occurred. Just because there is more data from which to work doesn't mean the aforementioned problems of representation and unequal power relations embodied in the data are resolved. The intimate intermingling of digital and material facets of life means that the production of geosocial media content is often strongly connected to place-based features and events, but also that longstanding inequalities and situational or contextual constraints distort the representativeness of such data sources. While we are sympathetic to such critiques of big data, we maintain that an explicitly geographical approach might be able to partially resolve some problems raised by earlier critiques. For example, while using geotagged tweets as one's sole data source might produce a flawed or incredibly partial analysis of an event like Hurricane Sandy, this data can also be used to answer broader questions around the geographies of the geoweb and how such spatialities might be conceptualized, as we do in this paper.

B. The Polymorphous Geographies of Social Media

Research into the geographies of social media has largely eschewed any explicit theorization of space and spatiality. Even where implicit, studies have tended to privilege a unitary understanding of space. For example, Takhteyev et al (2012) employ a networked or topological understanding of socio-spatial relations by focusing on social connections between Twitter users, while work by Goodchild and Li (2012) and Haklay (2010) focused on questions about the quality and locational accuracy of volunteered geographic information. Other similar work on mapping the user-generated and social media data from the geoweb has alternatively tended to over-emphasize the groundedness of such content in particular places, or how particular place-specific attributes, such as religion and language, are reflected in this data (Graham and Zook 2011; Shelton et al 2012; Graham and Zook 2013).

Yet conceptualizations of space that focus on only a single understanding of it necessarily belie the complexity of forms that socio-spatial relations take. In order to overcome this issue, Crampton et al (2013) have proposed a loose framework for going 'beyond the geotag' in analyzing geosocial media data. They argue that researchers should explicitly recognize the diversity of spatialities embodied in social media content in order to avoid over-privileging what amounts to a simplified spatial ontology of latitude and longitude coordinates. Analyses that fail to go beyond a simplified spatial ontology – e.g. simply plotting data points in Cartesian space – often overlook the range of quantitative and qualitative approaches that allow one to better understand the context and meaning of such big data, and tend to reinforce territorial or place-based dimensions of data at the expense of thinking space relationally (cf. Massey 1991; Amin 2002).

We use this constructive critique of earlier work on mapping user-generated data as a foundation for positioning our intervention within a pre-existing framework for understanding socio-spatial relations. Specifically, we adapt Jessop et al's (2008) TPSN framework in order to construct a more holistic picture of the variegated landscapes of the geoweb, emphasizing both the territorial and relational dimensions of this data. Jessop et al. argue that by focusing on the polymorphous nature of socio-spatial relations and their expression through the dimensions of territory, place, scale and networks, a more open and comprehensive understanding of socio-spatial relations is possible. They note that most socio-spatial research is concerned with just one of these dimensions, committing what they refer to as the fallacy of 'one-dimensionalism'.

Instead, Jessop et al offer the TPSN framework as a kind of metatheory to emphasize the complex and variable nature of socio-spatial relations as simultaneously and variably bounded and coherent (territory), as differentiated and embedded in particular contexts (place), as hierarchically organized (scale) and as interconnected or interdependent (networks). Each of these dimensions must be understood as always co-present and interconnected with the others; they can only be separated analytically, but never in practice, as, for example even the most global phenomena are always grounded in particular experiences of place, and vice versa. This approach avoids privileging any single dimension of space and instead highlights the ways that the technologies and knowledges of the geoweb and social media are expressed in a number of different ways simultaneously. It is for this reason that our empirical analysis, taken up in Section 4, does not separate out each of these dimensions when considering different ways of looking at the data. While some dimensions are more prevalent in a given representation than others, no representation is illustrative of just a single dimension.

As the TPSN theoretical framework is a key part of our analysis, further explanation of each of the four dimensions of socio-spatial relations is warranted. While some of the specifics of Jessop et al's explanation of TPSN may not be especially relevant in the case of Hurricane Sandy (for instance, this paper does not focus on divisions of labor or nongovernmental international regimes which are important in the context of political-economic analysis that the framework was originally intended for) their framework offers a useful heuristic for thinking about the multiple spatialities of social media data, or any social phenomenon for that matter. Table 1 outlines their original conceptions of each of the four dimensions, as well as our adaptations of these ideas to the context of analyzing Hurricane Sandy's data shadows.

Table 3.1: Operationalizing the TPSN Framework

Dimension	As articulated by Jessop et al (2008)	Our operationalization
Territory	Bordering, bounding, parcelization, enclosure	Locality, proximity, materiality
Place	Proximity, spatial embedding, areal differentiation	Lived experiences, individual perceptions
Scale	Hierarchization, vertical differentiation	Hierarchical organization, 'size' of areal lens
Networks	Interconnectivity, interdependence, rhizomatic differentiation	Interconnectivity, non-proximate, relational space

We employ the concept of *territory* to understand how user-generated content is spatialized in particular localities through the mirroring of offline, material phenomena occurring there. While resembling a conventional definition of the spatiality of big data as simply a set of latitude and longitude coordinates, it more importantly provides insight on the general contours, and occasional discrepancies, between our understandings of the materiality of a given phenomena and its online reflection. Such a connection to particular localities is tempered by integrating a focus on *scale*. While scale is a slippery concept – varying in meaning depending on whether one is using the concept in the context of an urban political economy, physical geography or GIScience approach (just to name a few competing understandings), our attention rests on the ways that using different scalar constructs, such as the juridical boundaries of neighborhoods, zip codes, census tracts, cities, states, and so on, can alter perceptions of the socio-spatial processes embodied in these data shadows.

In addition to territory and scale we also integrate a focus on *place*, or the lived dimension expressed in the qualitative information contained within these datasets. Rather than assuming a simple relationship between a piece of social media content and the location to which it is tagged, we work to understand the significance of these localities to the users producing such representations and the social contexts in which such content is embedded, acknowledging the potential for these experiences to be spatially distanciated from the locations in which a given event occurs. For this reason, we shift the notion or theme of 'proximity' from place, as it is conceived by Jessop et al, to territory, as mentioned above, to preserve an understanding of place that is more closely aligned with conventional understandings within geographic thought (Cresswell 2004). Finally, we turn to connecting these lived experiences of place to the broader patterns evident in territorial and scalar frameworks, through a focus on socio-spatial *networks*, or relational spaces. That is, understanding territories, places and scales as bounded or limited ignores the connections between localities, and the ways that social processes are increasingly extensive over long distances. In short, the network dimension reflects that one cannot fully understand the geographies of place-based phenomena without understanding that place's connections to other localities.

Rather than simply gathering such data, aggregating them and then displaying their location on a map, the TPSN approach provides a richer set of socio-spatial dimensions that can be used to understand the production and consumption of geographically referenced big data such as that which is derived from social media.

Utilizing the TPSN framework also provides an important connection between research on the geoweb and big data to the broader field of geography and socio-spatial theory. It allows us to clearly demonstrate that the socio-spatial relations of geosocial media are not divorced from socio-spatial relations more generally, and ultimately helps illuminate the full range of human experiences that are evident in such data shadows.

III. Collecting and Analyzing Big Data from Social Media

In order to operationalize the TPSN framework in the context of big data, this paper analyzes the data shadows of Hurricane Sandy through a specially designed software program that collects all geocoded tweets worldwide through the Twitter API, or application programming interface. While websites and social media platforms often provide APIs as part of a business strategy, researchers have begun to take advantage of these tools to access the significant amounts of data being generated through such platforms. The specially designed software used here was already operating prior to Hurricane Sandy, allowing us to select only tweets sent from within the United States between October 24 and October 31, 2012 that contain the keywords 'sandy', 'frankenstorm'[7], 'flood', or variations thereof. This results in a dataset consisting of 141,909 tweets. While each tweet has a variety of associated metadata, ranging from the actual tweet text to the number of friends that that Twitter user has, this study only uses the actual text, the timestamp and the location of the tweet.

---

[7] The term "Frankenstorm" was widely used to refer to the landfall of Hurricane Sandy in the northeastern United States in late October 2012. The term was adopted both because of the intensity of the storm and its timing immediately before Halloween.

It should be noted that these 141,909 geotagged tweets represent only a fairly small percentage of the total number of Sandy-related tweets during this time period, as only approximately 1.7 percent of all tweets contain explicit geographic information. While techniques exist to derive locational information from user-provided location information in profiles, this introduces its own set of issues surrounding self-reporting, precision, geocoding accuracy and the difference between a user's home location and the location from where a particular tweet is sent[8]. It is for this reason that we focus only on the relatively clean dataset of tweets that contain explicit geographic information. But even within this dataset, there exists variation in how location is derived. Of our dataset, 82 percent of the geotagged tweets contain an actual latitude/longitude coordinate pair, derived from the GPS sensor on a smartphone or through cell tower triangulation. The other 18 percent only contains a 'place' specification, which can vary in precision from the country level to cities to neighborhoods or points of interest. For obvious reasons, tweets that only have higher-level place information are filtered out when doing a local level analysis (e.g., tweets with only city-level definitions must be discarded when doing a neighborhood-level analysis).

Given the relatively large dataset – thousands of points – one must be mindful of three significant challenges, that if not dealt with correctly can prevent even the relatively straightforward exercise of mapping points in Cartesian space from yielding

[8] See Stephens and Poorthuis (forthcoming), who were able to find location data for 25% of all users, and Graham et al (2014), who show that geocoding accuracy varies substantially based on both location and language, for more discussion of these issues.

useful insights. First is the issue of overplotting. Plotting thousands of points onto a single map makes it difficult to distinguish between the intensity or size of different clusters. Second, regardless of the phenomena under study, places that are already large content producers will almost certainly produce high amounts (in absolute terms) of social media references to the phenomenon of interest. The third, and related, challenge is that the uneven spatial distributions of tweets means the amount located in any one region varies considerably, affecting the confidence with which we can infer differences from location to location.

To overcome these three challenges we use an approach that overlays the area of study with a grid of hexagonal cells of varying size. We use hexagonal cells instead of the more common rectangular grid cells for two specific reasons. First and foremost, cartographically, hexagons make it easier to increase the size of each cell (thus negating the use of smoothing, which is not always a good practice when dealing with phenomena that are not necessarily 'smooth') while still allowing the reader to discern contours. Square cells, as opposed to hexagons, are much more distracting to map readers and thus make it more difficult to determine the spatial pattern of a phenomenon (Carr et al 1992). Second, hexagons also have a higher representational accuracy (Scott 1985) and, when used in statistical analysis share a direct boundary with 6 neighbors, instead of the 4 direct neighbors of squares. Being able to vary the size of the cells allows us to use 'appropriate' cell sizes for different scale levels as well as address the potential effect of the Modifiable Areal Unit Problem (cf. Poorthuis and Zook 2015, for a more detailed discussion of this approach). In this paper, we use 65-kilometer wide cells for the national scale and 2-kilometer wide cells for the urban scale – both chosen to balance the generic with the particular so the map reader can distinguish larger patterns while not losing some smaller idiosyncrasies. Furthermore, we use a sample of 138,021 random tweets sent from the United States during the same time period from which our database of Sandy-related tweets was drawn in order to normalize data within these hexagonal units. Although population is often used for normalization purposes, using a random sample of tweets allows us to normalize by 'Twitter population' instead. The sample is drawn from the same proprietary system as the Sandy dataset, which allows for the extraction of random samples of all geotagged tweets of any size. In this case, we have chosen the sample to be roughly the same size as the dataset under study. We calculate both the number of Sandy-related tweets as well as the number of 'random' tweets. We then use both counts to calculate a variation on the odds ratio, referred to as location quotient in spatial economics, taking the lower bound of the 99.9 percent confidence interval for each cell as follows:

$$OR_{lower} = e^{\ln(OR_i) - 3.29 * \sqrt{\frac{1}{p_i} + \frac{1}{p} + \frac{1}{r_i} + \frac{1}{r}}}$$

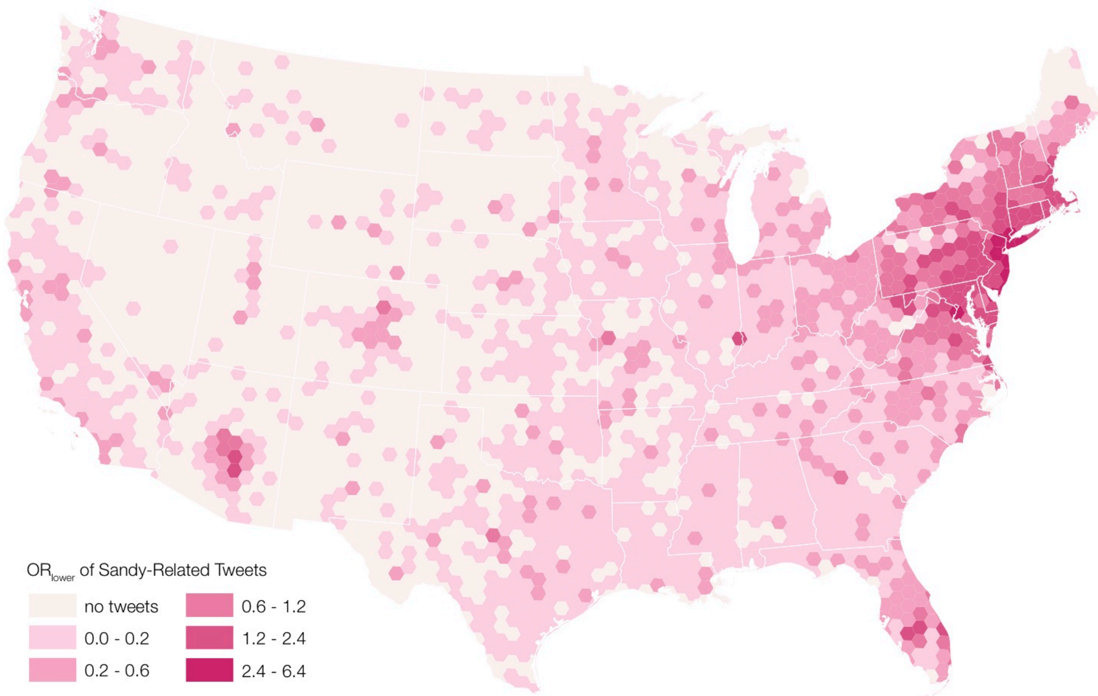
Where p_i is the number of tweets in hexagon i related to the phenomenon of interest and p is the sum of all tweets related to the phenomenon; r_i is the number of random tweets in hexagon i and r the sum of all random tweets. This results in a ratio where a value of 1 means that there are exactly as many data points for the phenomenon as one would expect based on the random sample. An odds ratio greater than 1 means that we can say, with 99.9 percent confidence, that there are more points related to the phenomenon than one should expect, and vice versa for anything under 1.

Although the entire dataset contains more than 3 billion tweets as of August 2013, the case studies in this paper only use a subset of this data based on the query outlined previously. It is important to highlight that we cannot draw direct correlations between the size of our datasets and the veracity of insights that can be drawn from those data. Although these data offer the raw materials for analysis and understanding, simply plotting points on a map is an insufficient endeavor to comprehend the polymorphous and variegated geographies of social media as conceptualized using the TPSN framework. As such, we will augment a more quantitative and GIS-oriented analysis with a qualitative analysis of the content of tweets. Such analysis is not a significant departure from longstanding traditions of cultural landscape interpretation within geography, though the landscapes that we interpret here are the digital representations of material actions, patterns, and processes, or what have previously been referred to as 'cyberscapes' (Crutcher and Zook 2009; Graham and Zook 2011; Shelton et al 2012). This paper's methodological approach is thus necessarily interlinked with the conceptual approach of the TPSN framework.

IV. Socio-spatial Dimensions of Hurricane Sandy's Data Shadows

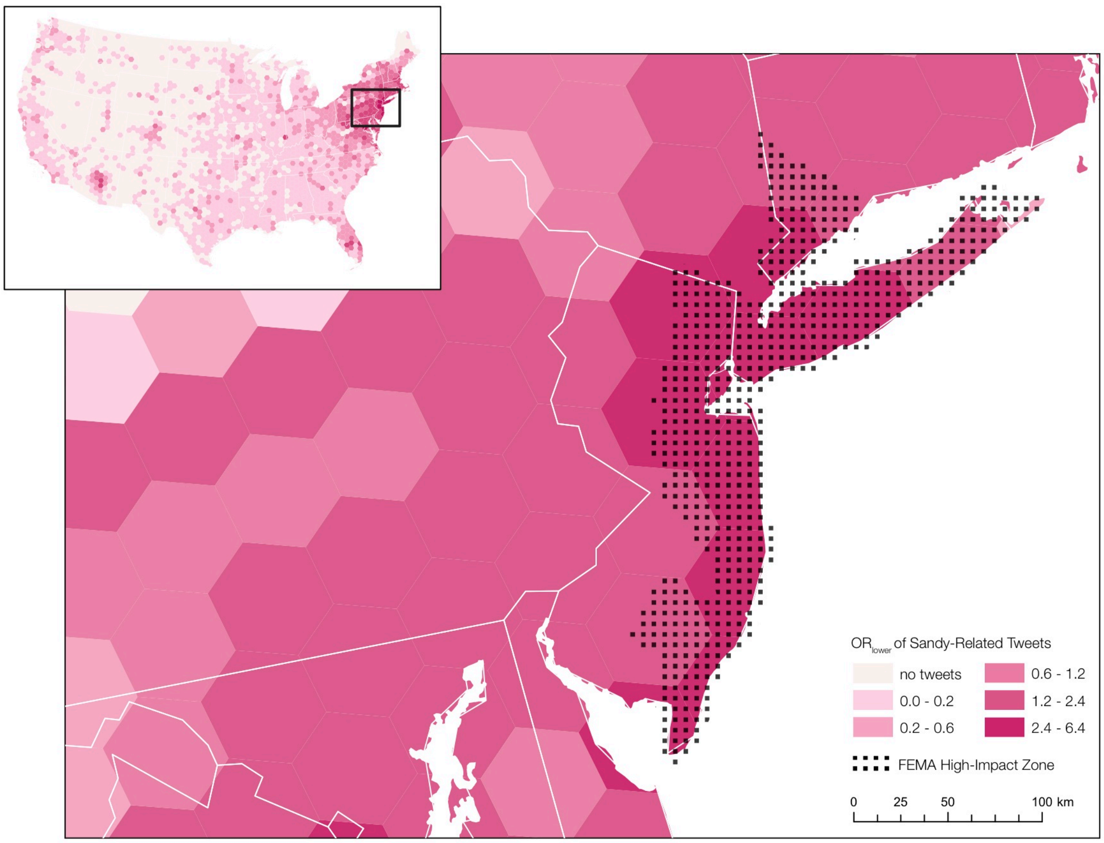
In order to better understand the diversity of ways that social media data shadows reveal or conceal useful information, we now turn to interrogating the aforementioned dataset of tweets related to Hurricane Sandy through the four core dimensions of socio-spatial relations – territory, place, scale and networks – as outlined by Jessop et al (2008). While each of the visualizations might be loosely placed under one of the four headings, we have intentionally chosen not to present them separately, so as to emphasize that each visualization demonstrates the fundamentally multiplicitous socio-spatial relationships of the geoweb.

Figure 3.1: Sandy-related Tweets Across the United States



The first, and most obvious, way to approach these data is to look at the distribution of Sandy-related tweets at a broad spatial scale, in this case looking at the continental United States. Using the odds ratio metric explained in the previous section, Figure 3.1 clearly shows a significant concentration of Sandy-related tweets along the eastern seaboard of the US, especially in those places that were most affected by the storm, with approximately 30 percent of all Sandy-related tweets being located in the New York City metropolitan area. While there are some intriguing anomalies, for instance the cluster of tweeting around Phoenix, Arizona, this map is largely unsurprising given the material manifestation of Hurricane Sandy in the Northeastern US.

Figure 3.2: Sandy-related Tweets along Eastern Seaboard



Zooming in to the affected area, there appears to be important utility in employing social media data to measure the extent of storm damage (see Figure 3.2). Using the same data as Figure 3.1, this map adds a layer representing the official 'High-Impact Zone' as determined by the Federal Emergency Management Authority (FEMA), which is roughly congruent with the areas with the highest relative amounts of Sandy-related tweeting activity. This connection is further bolstered by the fact that the New York metropolitan area suffered the greatest financial losses from the storm, totaling approximately $19 billion (Gormley 2012).

To be clear, this map is not intended to discount that other populated areas, such parts of Pennsylvania, Virginia, and the Caribbean (which we've excluded altogether from this analysis) were also hit hard by Sandy. Rather, we use the 'High Impact Zone' definition in Figure 3.2, to demonstrate a clear connection to the places in which that content was produced, underlining the territoriality of geosocial media data. But highlighting this territoriality is merely the first step of the analysis. In order to place the groundedness of this content in context, we must also examine how it is intertwined with other dimensions of socio-spatial relations.

For example, despite the overall devastation experienced by New York City and the surrounding areas, it is problematic to assume that New York City as a place is entirely coherent and that people's experiences of the storm were uniform throughout different areas of the city. By integrating a focus on scale with our already established focus on territory, we can get a better idea of the actual contours of Sandy-related tweeting in New York City (see Figure 3.3).

When taking a closer look at New York City, we can adjust the size of the hexagonal cells used to aggregate tweets, which in turn creates a finer grained surface for analysis. While we are still examining the territoriality of tweets, we have also in this moment shifted scales, essentially disaggregating the coarser definition of the New York metropolitan area used in Figures 3.1 and 3.2 into a series of smaller spatial units to allow for intra-urban analysis. Figure 3.3a highlights (via text call outs in the maps) places in the city where significant events during the storm coincide with higher-than-average levels of tweeting, while Figure 3.3b highlights places where major events were reported by the media but had relatively few tweets.

<u>Figure 3.3: Sandy-related Tweets in New York City Metropolitan Area</u>

Figure 3.3a demonstrates that a number of places that experienced significant damage were also major producers of Sandy-related tweets. Some areas with significant tweeting activity, such as the Lower East Side, which experienced significant flooding and power outages, are relatively wealthy, and even some poorer areas, such as Coney Island, had significant levels of tweeting activity. At the same time, however, some of the hardest hit places also had relatively little tweeting activity (see Figure 3.3b). For example, in Breezy Point, a fire destroyed more than eighty homes, but only a handful

of tweets come from that location. Sandy inflicted similar damage on large parts of the Rockaway Peninsula with very little mention in these places on Twitter. We are also able to see a general lack of tweeting from Staten Island, which has the unfortunate distinction of having nearly half of the Sandy-related deaths within the city, not to mention massive amounts of property damage in the Oakwood area. While some residents in these areas were likely preoccupied with more pressing matters than tweeting, this runs counter to examples in Figure 3.3a where significant amounts of tweeting correlated with high-damage locations. The differences between these two figures suggest that places on the spatial periphery of the metropolitan area, e.g., Staten Island or the Bronx, are more likely to be marginalized within data shadows than more central locations, e.g., Manhattan and Brooklyn. While there is no definitive explanation for these discrepancies between damage and tweeting activity, it is above all demonstrative of the fact that the correlation between these variables changes across scales, thus necessitating the inclusion of the scalar dimension in any similar analysis.

Thus, shifting the analysis from the national to the urban scale reveals that the relatively strong correlation between tweet density and territories most affected by Sandy breaks down at finer scales of analysis; a finding that raises concerns about some of the practical applications of mapping geosocial media in disaster situations. In other words, strategies that rely upon the data shadows of social media for determining the allocation of scarce resources in a crisis need to consider the biases and permutations that accompany the production of this data.

Figure 3.4: Tweets about the 57th Street Crane in New York City

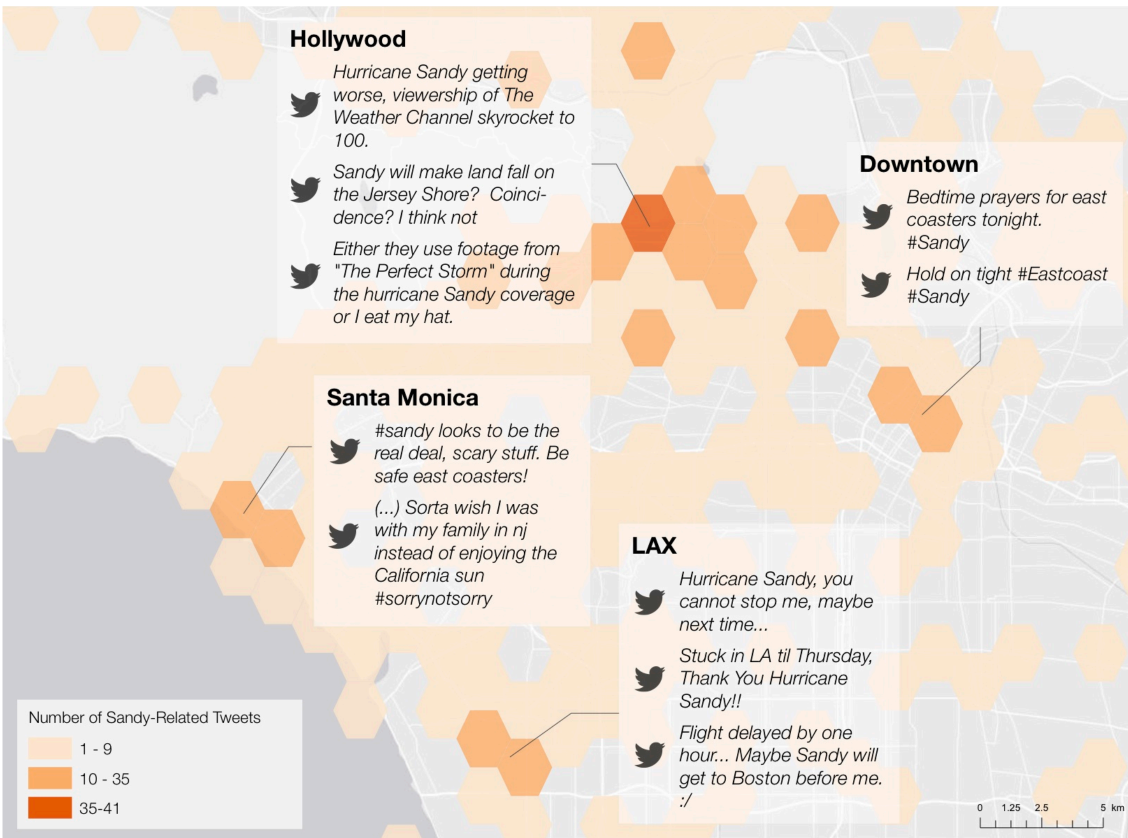
For this reason we argue for the utility in proceeding with an iterative analysis that focuses on specific events, rather than simple mappings of terms like "sandy", "frankenstorm" and "flood." For instance, mapping the location of the 774 tweets mentioning "crane" during the storm, we are able to pinpoint the location of the now infamous 57th Street crane that was left dangling in the aftermath of the storm (see Figure 3.4). Although we are cautious about the potentials of automated, algorithmic analysis of big data in many contexts, this example highlights the potential of such analysis in places characterised by thick data shadows, such that a kind 'early detection' mechanism might be able to automatically identify spatial and temporal irregularities in the data.

The 57th Street crane example demonstrates the value in extending our analysis beyond the territorial and scalar dimensions and into the lived dimension of place-making. As useful as it is to use these virtual expressions of material phenomena to locate these events in Cartesian space, stopping there neglects the way that these data are reflective of particular experiences of place by particular individuals. This is, we assert, a necessary, but thus far largely overlooked, contribution that geographers can make to the broader study of social media activity. A focus on the qualitative experience of place embodied in this data and resulting data shadows offers a significant opportunity for geographers and others interested in the spatial dimensions of social media, and can create a much more nuanced understanding of these dimensions when paired with the more general analysis of territory and scale emphasized in Figures 3.1-3.3. While this may have more to do with the post hoc analysis of such catastrophes than for the immediate disaster response, it highlights the importance of attending to the qualitative information and social context of such data, even during disaster response, and not over-privileging automated systems for sentiment analysis, which leave significant potential for misinterpretation.

Another important consideration is that a focus merely on the greatest concentrations of tweeting activity provides relatively little insight into the array of meanings encoded into social media datasets. While we can use the first slice of the territorial dimension to understand the basic spatial distribution of tweeting activity (as evidenced in Figures 3.1-3.3), this assumes a level of homogeneity within the qualitative information contained within the tweets themselves. It is similarly important to consider that places which may not have especially high concentrations of tweeting activity, and which might be quite far from those places which do, also have something to tell us about the spatiality of social media. For instance, of the nearly 142,000-geotagged tweets used in Figure 3.1, only 42,000 or so of those are in the New York metropolitan area. So what are we to make of the remaining 100,000 tweets if we focus only on those places with the most activity? Indeed, what is the utility of 'big data' if we're ignoring such a significant portion of it?

One corrective to this, inspired by Doreen Massey's idea of a global sense of place (Massey 1991), is to turn our attention to a greater diversity of places, including those with relatively few Sandy-related tweets and those quite far from New York in absolute distance, but actually quite proximate in relational terms. By combining a focus on the place and network dimensions of socio-spatial relations as outlined in the TPSN framework, we can begin to put a greater emphasis on understanding the totality of the dataset. For example, looking at Sandy-related tweets in the Los Angeles metropolitan area, of which there are only 2,476, one sees a number of revealing interand intra-urban geographies (see Figure 3.5). Although Los Angeles as a whole was thousands of miles away from the physical manifestations of Hurricane Sandy, the data shadows produced by Twitter users in different parts of this metropolitan area vary considerably from each other.

Figure 3.5: Sandy-related Tweets in the Los Angeles Metropolitan Area



With New York City and Los Angeles being the archetypical 'global cities' of the United States, we know that one of the most important linkages between them is by way of air travel (Derudder et al 2007), so a cluster of people in each city's airports concerned about their ability to fly cross-country with the impending storm is perhaps unsurprising. But a variety of relational connections are visible in other locales as well, even in the absence of similarly obvious rationales. For instance, though Hollywood has the greatest overall number of Sandy-related tweets in Los Angeles, tweets originating from this area seem filtered through a film and television-centric framing specific to the area, with references linking Sandy to films like *Judgement Day* and *The Perfect Storm*, the reality television show *The Jersey Shore* and The Weather Channel's tendency to dramatize weather events in order to promote their own programming. Read through the lens of TPSN, understanding Los Angeles' place within the broader landscape of Sandy-related tweeting, as well as its internal variegation, brings to the forefront the dimensions of place and networks, or the "presence of both the proximate and the remote at the same geographical level" (Amin 2002: 389).

While Los Angeles doesn't necessarily have a particularly prominent place in the territoriality of Sandy-related tweeting at any scale, these examples highlight the utility of going beyond just looking at those areas with the highest concentrations of tweeting activity. Instead, a closer reading of social media content in a variety of locations reveals how spaces that might otherwise be neglected in such analyses still provide important insights into the geographies of big data. Rather than simply matching information mediated by social media platforms to spatial locations, significant meaning can also be drawn from the interconnectivity and interdependence of those data, raising the question of what the topology of connections between information producers and information itself tells us about these material phenomena. In other words, we can see not only where something happens in physical space, but how an event connects to other spaces both near and far through network ties. Multidimensional understandings of socio-spatial processes are important across a range of issues, beyond just our readings of such processes through big data. For instance, natural disasters like Hurricane Sandy are conventionally understood as having very particular localized effects around the areas most affected, in this case the New York metropolitan region and the eastern seaboard. But any understanding of Sandy as being spatially delimited to these places would be lacking. Equally important is how those impacts were shaped but by more complex social forces that stretch beyond these localities but nevertheless structure experiences of and responses to such events, such as the genesis of climate change and its impacts on increasingly irregular and volatile weather patterns or the political-economic structures that cause predominantly poor and minority neighborhoods to be the most vulnerable to such disaster events (cf. Smith 2006). But such logics also apply to the ways that disasters or other such events reach beyond these localities in their effects, which are further mediated by place-based experiences in those locations, as is demonstrated by the different experiences and interpretations of Hurricane Sandy within Los Angeles as seen in the Twitter data.

Although the relational connection between New York City and Los Angeles makes for a convenient example, it is less clear how socio-spatial networks across the US were articulated through social media during Hurricane Sandy. If the tweeting from

LAX during Sandy is indicative of a broader pattern induced by airplane-enabled connectivity, can we find similar connections to other locations across the US as well? The goal of such an analysis is to demonstrate the extent of the relational dimension of social media activity beyond the obvious connections between global cities such as New York and Los Angeles.

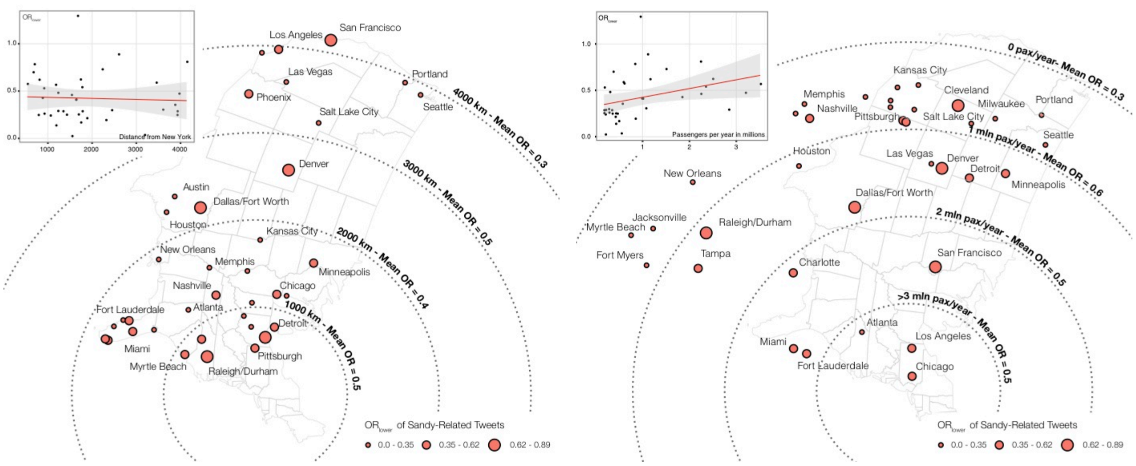
Using T-100 Domestic Market data from the Research and Innovative Technology Administration (RITA) on flights and the number of passengers between city pairs in 2012, we determined the 50 cities that have the most passenger traffic with New York City, ranging from Chicago (3.5 million passengers back and forth) to Kansas City (175,000 passengers). Since operations and activities at some airports close to New York were directly affected by Sandy's landfall, we exclude any airport within 500 kilometers of Manhattan in this analysis. For the remaining airports we used a buffer of 5 kilometers to collect all Hurricane Sandy related tweets and calculated the lower bound of the odds ratio as we did for the hexagonal cells used in Figures 3.1 to 3.3. If relational networks did not play a significant role in Sandy-related tweeting, one would expect to see a direct distance decay effect: as the distance from New York City increases the odds ratio should decrease.

Figure 3.6a, however, shows that physical distance has no significant relationship with the relative level of tweeting activity about Hurricane Sandy as is evidenced by both the scatterplot[9] and the map (Spearman's rho is -0.05). The map uses

[9] The red line through the scatterplot indicates a fit using a linear model. Confidence interval of the fit is indicated in light grey.

an azimuthal equidistant projection with New York City as the center, where the size of each airport is proportional to its odds ratio. Airports that are equally distant in physical terms from New York have widely diverging measures of Sandy-related Twitter activity. In addition, the average odds ratio in each 1,000km zone does not decrease the further away one travels from New York.

Figure 6: Sandy's Socio-spatial Networks



In contrast, Figure 3.6b shows that the relational ties between each city and New York, measured by number of passengers, exhibits a much stronger positive correlation with the odds ratio metric of Twitter activity (Spearman's rho is 0.34). This figure preserves the directional bearing of each city with respect to New York City, but instead uses an inverse of the number of passengers to recalculate the relational distance between the cities. Airports are thus no longer displayed according to their physical distance from New York City, but rather based on the amount of passenger traffic between the two cities. Since the bearing has remained the same, airports with a higher intensity will move closer to New York along that line, and vice versa. In addition to the correlation coefficient, we can also visually determine that cities with a lower odds-ratio, such as Pittsburgh and Memphis, have a tendency to move towards the outer circles while cities with a higher odds ratio, such as San Francisco and Los Angeles, move relatively closer.

In other words, it is the relational connection to New York, measured by number of air travellers, not physical distance, which better explains the level of concern with Hurricane Sandy. This concern, however, can vary within metropolitan territories as evidenced by Figure 3.5 depending upon the scale of analysis; some parts of an urban area may have much stronger relational ties to distant cities, while other parts are largely disconnected from such translocal flows.

To test the extent to which the data shadows of Sandy-related tweeting are a localized phenomenon within certain parts of metropolitan areas (rather than a more generalized territorial phenomenon), we increased the initial buffer around each airport from 5 kilometres to 25 kilometers. Thus, rather than just capturing neighborhoods that are spatially proximate to the airport, this measure captures a much wider swath of each metropolitan area. In the case of Los Angeles, this includes the entirety of the territory shown in Figure 3.5 and beyond. With this larger buffer, there is a near-reversal of the correlations illustrated in Figure 3.6, as Pearson's rho for total number of passengers is now 0.06 (rather than 0.34), while the distance effect starts to emerge (rho is -0.15). In other words, even though the socio-spatiality of a phenomenon like Sandy is expressed partly through a network of connections between territories, these connections are very much bounded by the locally-specific practices of place. So not only can we discern more complex socio-spatial relations than just the immediate experience of a natural disaster through this data, but we can also understand how the spatially distanciated networks have their own territorial groundings, just not only in those places one might expect. This once again highlights the complex ways in which the digital data shadows of a material event are manifest through the intertwining of different dimensions of social space.

As evidenced by each of these examples, Sandy's data shadows are not evenly distributed through the continental United States. They are instead quite intense in some locations, while hardly reaching other physically adjacent sites. Airline passenger movement partially explains how the data shadows of social processes are stretched over physical space and user-generated social media provides another indicator for better understanding the production of these relations. Disasters do not transpire in a single, unitary and bounded locale, but are embedded within complex and evolving socio-spatial relations that stretch unevenly across space. Some places are connected quite closely because of their political and economic interdependencies or dense social ties. Other places, while physically closer, which lack such substantive relational connections tend to have quite different experiences of such events.

V. Conclusion

The analysis of the data shadows of Hurricane Sandy presented here reveals relatively few surprises. Tweeting was largely concentrated in the areas hit hardest by the hurricane, with more distant areas having many fewer Sandy-related tweets. This analysis, however, has expanded via more holistic methodological and conceptual approaches, allowing us to demonstrate the shortcomings of simply plotting points on the earth's surface and assuming a one-to-one relationship between the location of tweets and the material events about which they are created. This kind of commonplace approach fails to acknowledge the unevenness of tweeting at different scales, it ignores the full range of knowledges represented in the content of tweets which themselves are locally specific, and it overlooks the spatially-distanciated, relational networks which complicate any assumptions of a uniform distance decay effect, among other things.

While there is undoubted potential in using social media in times of crisis, we worry that too much of the discourse and practice of crisis mapping, let alone other applications of this kind of data, relies on the relatively simple spatial ontologies and epistemologies that we have critiqued here. That is, seeing spatial concentrations of social media activity in disaster situations as being equivalent to areas in need of relief vastly oversimplifies the ways that social media is used in disaster situations, while also potentially reinforcing offline social inequalities by failing to provide relief to areas which may not be producing such content because of lack of access to the appropriate technologies or material conditions preventing the use of such tools (e.g., power outages). Geosocial media data can undoubtedly provide an important window into understanding disaster situations and formulating responses to them, but we would argue that any utilization of this data would be wise to account for the complexities that it embodies. While this need for problematization and caution might limit the usefulness of this kind of data in the immediate aftermath of disasters, disaster response is often a long and laborious process (as has been the case with Hurricane Sandy), so it is important to note that this data may well remain useful for analysis after the crisis event itself.

The promise of utilizing such big data sources for social scientific analysis isn't solely in the size of the dataset, but the wealth of social processes that are encoded in such data. Thus, even though our case study of Hurricane Sandy doesn't present any radically new empirical insights into the geography of Twitter, or user-generated geographic information more broadly, we believe that this case study has allowed us to articulate three key conceptual and methodological points that should inform any similar analyses of geosocial media data in the future.

First, we have shown the utility of using small subsets of big data sources for social and spatial analysis. Starting with a large archive exceeding 3 billion geotagged tweets, we used only roughly 140,000 Sandy-related tweets for this case study. So even while this might have meant that there were just a few dozen data points in a given neighborhood in some cases, this amount of data is more than sufficient to gain statistically significant insights from our quantitative analysis, while also making qualitative analysis of these tweets more manageable. It is important to again emphasize that more data doesn't necessarily lead to more meaningful results, or a more accurate depiction of the world around us, something generally obscured by the contemporary fetish for ‘bigness’ in data.

Second, we have shown the importance of a mixed methods approach to understanding big data. A quantitative mapping of tweet density, however technically sophisticated, ultimately stops short of understanding the complex and polymorphous geographies of such data without also performing a qualitative analysis of the actual tweets and the context in which they are produced, or even employing a diversity of quantitative methods, such as social network analysis. Similarly, a qualitative analysis of such big data sources, even when narrowed down to just 140,000 data points, is impractical, if not impossible, without some a priori analysis and filtering based on quantitative methods. Such a mixed method approach not only avoids a kind of naïve empiricism with respect to big data that is currently prevalent, but also fundamentally points towards big data as embodying a variety of social and spatial relations which can begin to parse out through such analysis.

Finally, we have argued for the value of employing existing conceptual frameworks, such as Jessop et al’s TPSN framework, to better understand the complexities of user-generated content and the socio-spatial relations they embody. While most of the existing work on the geoweb has failed to explicitly theorize socio-spatial relations, we have used the case of Hurricane Sandy and its data shadows to demonstrate the utility of the TPSN framework and its underlying analytical dimensions to produce much deeper understandings of space and spatiality as embodied in user-generated geographic information. We believe that these three contributions can help to provide a firmer foundation for future analyses of geosocial media data, highlighting the complex and variegated socio-spatial relations represented in such data sources.

VI. References

Amin, Ash. 2002. "Spatialities of Globalisation." *Environment and Planning A* 34 (3): 385–399.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine* 15(7).

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–679.

Cairncross, Frances. 1997. *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Harvard Business Press.

Carr, Daniel B., Anthony R. Olsen, and Denis White. 1992. "Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data." *Cartography and Geographic Information Systems* 19 (4): 228–236.

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb." *Cartography and Geographic Information Science* 40 (2): 130–139.

Crawford, Kate. 2013. "The Hidden Biases in Big Data." *Harvard Business Review*. 1 April. Available from: http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html.

Cresswell, Tim. 2004. *Place: A Short Introduction.* Blackwell.

Crutcher, Michael, and Matthew Zook. 2009. "Placemarks and Waterlines: Racialized Cyberscapes in Post-Katrina Google Earth." *Geoforum* 40 (4): 523–534.

Derudder, Ben, Frank Witlox, and Peter Taylor. 2007. "U.S. Cities in the World City Network: Comparing their Positions using Global Origins and Destinations of Airline Passengers." *Urban Geography* 28(1): 74–91.

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

Elwood, Sarah, Michael F. Goodchild, and Daniel Z. Sui. 2012. “Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice.” *Annals of the Association of American Geographers* 102 (3): 571–590.

Floridi, Luciano. 2012. “Big Data and Their Epistemological Challenge.” *Philosophy & Technology* 25 (4): 435–437.

Goodchild, Michael. 2007. “Citizens as Sensors: The World of Volunteered Geography.” *GeoJournal* 69 (4): 211–221.

Goodchild, Michael F., and J. Alan Glennon. 2010. “Crowdsourcing geographic information for disaster response: a research frontier”. *International Journal of Digital Earth* 3 (3): 231-241.

Goodchild, Michael, and Linna Li. 2012. “Assuring the quality of volunteered geographic information.” *Spatial Statistics* 1 (1): 110–120.

Gormley, Michael. 2012. “Cuomo: Sandy Cost NY, NYC $32b in Damage and Loss”. *The Associated Press.* 26 November. Available from: http://bigstory.ap.org/article/cuomo-bloomberg-brief-ny-delegation-storm-cost

Graham, Mark. 2012. “Big data and the end of theory?” *The Guardian*. 9 March. Available from: http://www.guardian.co.uk/news/datablog/2012/mar/09/big-data-theory.

Graham, Mark. 2013. “The Virtual Dimension”. In *Global City Challenges: Debating a Concept, Improving the Practice*, edited by Michele Acuto and Wendy Steele. Palgrave. pp. 117-139.

Graham, Mark, Scott A. Hale, and Devin Gaffney. 2014. “Where in the World Are You? Geolocation and Language Identification in Twitter.” *The Professional Geographer* 66 (4): 568–578.

Graham, Mark, and Taylor Shelton. 2013. “Geography and the Future of Big Data, Big Data and the Future of Geography.” *Dialogues in Human Geography* 3 (3): 255–261.

Graham, Mark, and Matthew Zook. 2011. “Visualizing Global Cyberscapes: Mapping User-Generated Placemarks.” *Journal of Urban Technology* 18 (1): 115–132.

Graham, Mark, and Matthew Zook. 2013. “Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web.” *Environment and Planning A* 45 (1): 77–99.

Graham, Mark, Matthew Zook, and Andrew Boulton. 2013. “Augmented reality in urban places: contested content and the duplicity of code.” *Transactions of the Institute of British Geographers* 38 (3): 464–479.

Gregory, Sean. 2010. "Cory Booker: The Mayor of Twitter and Blizzard Superhero". *TIME Magazine.* 29 December. Available from: http://www.time.com/time/nation/article/0,8599,2039945,00.html

Haklay, Muki. 2010. "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets." *Environment and Planning B: Planning and Design* 37 (4): 682–703.

Haklay, Muki. 2012. "'Nobody wants to do council estates': digital divide, spatial justice and outliers". Paper presented at the 108th Annual Meeting of the Association of American Geographers. New York, NY. February 25, 2012.

Hey, Tony, and Anne Trefethen. 2003. "The Data Deluge: An e-Science Perspective". In *Grid Computing: Making the Global Infrastructure a Reality,* edited by Fran Berman, Geoffrey Fox, and Tony Hey. John Wiley and Sons. pp. 1–17.

Hey, Tony, Stewart Tansley, and Kristin Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

Issenberg, Sasha. 2012. *The Victory Lab: The Secret Science of Winning Campaigns*. Crown.

ITO World. 2010. "ITO World at TED 2010 – Project Haiti". *ITO World Blog.* February 12. http://itoworld.blogspot.com/2010/02/ito-world-at-ted-2010-project-haiti.html

Jessop, Bob, Neil Brenner, and Martin Jones. 2008. "Theorizing Sociospatial Relations." *Environment and Planning D: Society and Space* 26 (3): 389–401.

Kitchin, Rob. 2013. "Big Data and Human Geography: Opportunities, Challenges and Risks." *Dialogues in Human Geography* 3 (3): 262–267.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Lászlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323 (5915): 721–723.

Leetaru, Kalev. 2011. "Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space." *First Monday* 16 (9).

Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook. 2013. "Mapping the Global Twitter Heartbeat: The Geography of Twitter." *First Monday* 18 (5).

Lehrer, Jonah. 2010. "A Physicist Turns the City Into an Equation." *The New York Times*. 17 December. Available from: http://www.nytimes.com/2010/12/19/magazine/19Urban_West-t.html.

Leszczynski, Agnieszka. 2012. "Situating the Geoweb in Political Economy." *Progress in Human Geography* 36 (1): 72–89.

Li, Linna, and Michael F. Goodchild. 2010. “The role of social networks in emergency management: a research agenda.” *International Journal of Information Systems for Crisis Response and Management* 2 (4): 48-58.

Liu, Sophia B., and Leysia Palen. 2010. “The New Cartographers: Crisis Map Mashups and the Emergence of Neogeographic Practice.” *Cartography and Geographic Information Science* 37 (1): 69–90.

Lohr, Steve. 2012. “The Age of Big Data”. *The New York Times.* 11 February. Available from: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Manovich, Lev. 2011. "Trending: the promises and the challenges of big social data." In *Debates in the Digital Humanities*, edited by Matthew K. Gold. University of Minnesota Press. pp. 460-75.

Massey, Doreen. 1991. “A Global Sense of Place.” *Marxism Today* 35 (6): 24–29.

Mayer-Schonberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan/Houghton Mifflin Harcourt.

Miller, Harvey J. 2010. “The Data Avalanche Is Here. Shouldn’t We Be Digging?” *Journal of Regional Science* 50 (1): 181–201.

Poorthuis, Ate, and Matthew Zook. 2015. “Small Stories in Big Data: Gaining Insights From Large Spatial Point Pattern Datasets.” *Cityscape* 17 (1): 151–160.

Red Cross. n.d. **“**The American Red Cross and Dell Launch First-Of-Its-Kind Social Media Digital Operations Center for Humanitarian Relief”. Available from: http://www.redcross.org/news/press-release/The-American-Red-Cross-and-Dell-Launch-First-Of-Its-Kind-Social-Media-Digital-Operations-Center-for-Humanitarian-Relief

Roche, Stephane, Eliane Propeck-Zimmermann, and Boris Mericskay. 2013. “GeoWeb and crisis management: Issues and perspectives of volunteered geographic information.” *GeoJournal* 78 (1): 21-40.

Ruppert, Evelyn, John Law, and Mike Savage. 2013. “Reassembling Social Science Methods: The Challenge of Digital Devices.” *Theory, Culture & Society* 30 (4): 22–46.

Scott, David W. 1985. “Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions.” *The Annals of Statistics* 13 (3): 1024–1040.

Shelton, Taylor, Matthew Zook, and Mark Graham. 2012. “The Technology of Religion: Mapping Religious Cyberscapes.” *The Professional Geographer* 64 (4): 602–617.

Smith, Neil. 2006. "There's no such thing as a natural disaster." *Understanding Katrina: Perspectives from the Social Sciences*. Social Science Research Council. Available from: http://understandingkatrina.ssrc.org

Stephens, Monica, and Ate Poorthuis. Forthcoming. "Follow Thy Neighbor: Connecting the Social and the Spatial Networks on Twitter." *Computers, Environment and Urban Systems*.

Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. 2012. "Geography of Twitter Networks." *Social Networks* 34 (1): 73–81.

Torrens, Paul. 2010. "Geography and computational social science." *GeoJournal* 75 (2): 133–148.

Tsou, Ming-Hsiang, and Michael Leitner. 2013. "Visualization of Social Media: Seeing a Mirage or a Message?" *Cartography and Geographic Information Science* 40 (2): 55–60.

Twitter. n.d. "2012 Year on Twitter: Global town square". Available from: https://2012.twitter.com/en/global-town-square.html

Wald, David J., and James W. Dewey. 2005. "Did You Feel It? Citizens Contribute to Earthquake Science". *United States Geological Survey Fact Sheet* 2005-3016.

Wald, David J., Vincent Quitoriano, Lori A. Dengler, and James W. Dewey. 1999. "Utilization of the Internet for Rapid Community Intensity Maps." *Seismological Research Letters* 70 (6): 680–97.

Zook, Matthew, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." *World Medical & Health Policy* 2 (2): 7–33.

Zook, Matthew, Mark Graham, and Monica Stephens. 2013. "Data Shadows of an Underground Economy: Volunteered Geographic Information and the Economic Geographies of Marijuana". Unpublished Manuscript.

Chapter 4
Social Media and the City:
Rethinking Urban Socio-spatial Inequality Using User-generated Geographic Information

Published as:
Shelton, Taylor, Ate Poorthuis, and Matthew Zook. Forthcoming. "Social Media and the City: Rethinking Urban Socio-spatial Inequality Using User-generated Geographic Information." *Landscape and Urban Planning*.

I. Introduction

Attempts to make the ideas and practices of urban planning more 'scientific' through the application of new technologies have been persistent over the course of the last century (Fairfield 1994; Ford 1913; LeGates et al 2009; Light 2003). But as new sources of digital data – whether collected from mobile phones, social media feeds, sensors embedded in the built environment or any number of other sources – are increasingly able to be combined and cross-referenced to produce 'big data', there has been a revival of interest in mobilizing this data towards the end of a supposedly more holistic 'science of cities' (Bettencourt and West 2010; Batty 2012). The breadth of available data sources has expanded rapidly, allowing researchers to end their dependence upon official statistics on demographics, economic activity, traffic, and any number of other urban indicators.

But as these new data sources and new ways of approaching social science research have become more prominent, they have also faced increasing amounts of criticism. This is due in no small part to the hubris of big data advocates, as exemplified in Chris Anderson's (2008) now-infamous declaration of the 'end of theory'. And while claims to greater objectivity, neutrality, and accuracy are rampant among proponents of big data, boyd and Crawford (2012) astutely argue that these data are always the result of conscious, subjective decisions on the part of researchers, and are the result of inherently social processes. Indeed, it's important to keep in mind that in spite of the celebratory discourses around big data, many of these ideas and techniques have been around for considerable amounts of time (Barnes 2013; Graham and Shelton 2013). Wyly (2014), however, positions big data as the driver of a 'new quantitative revolution' in geography, a largely reductionist effort enabled by processes of neoliberalization which threaten the kind of situated research which geographers have become experts at producing. In describing what he sees as "the speedy pseudopositivism of tweet-space analysis", Wyly argues that "big data give us a quickly expanding, shallow view of the vast horizontal landscape of the desert of the present real, with each new technological advance accomplishing new kinds of devalorization of past generations of human knowledge" (Wyly 2014: 28).

While we are sympathetic to such critiques of big data, we also recognize that these traits are not inherent in the data themselves, nor in the analysis of such data. The use of big datasets isn't necessarily reductionist or ahistorical; these are, in fact, to echo Wyly's (2009) earlier analysis of quantification in geography, contingent circumstances. Indeed, we believe that big data can be quite easily fit into more critical-quantitative approaches to urban geography and planning (cf. Barnes 2009; Schwanen and Kwan 2009; Sheppard 2001; Wyly 2011). Though issues around the over-valorization of this kind of data remain, including how they might displace other forms of official statistical knowledge, we believe that there is also significant potential. For example, the finer spatial and temporal scale of these kinds of datasets provides a way to ask different kinds of questions than is possible with, for instance, Census data, which is often several years old by the time of its release, and is generally associated only with one's place of residence, and then aggregated to more-or-less arbitrary spatial units. As such, this paper highlights the potential in mobilizing big data sources for understanding urban socio-spatial processes, so long as such research is also explicit in its engagement with the appropriate conceptual and methodological frameworks, and built on a critical and contextualized understanding of the underlying data. When coupled with exactly the kind of historical and geographical context that Wyly sees missing from many big data analyses, we argue that these approaches can provide useful insight for urban planning and geographical research.

To this end, we use a dataset of geotagged tweets from Louisville, Kentucky to explore longstanding problems of socio-spatial inequality in the city. Louisville represents an interesting case study for a number of reasons: first, Louisville is something of an 'ordinary' city, especially when it comes to its reflection in these kinds of big data sources. Louisville is fairly average in the density of its social media footprint, meaning that the methods demonstrated in this paper are likely to be applicable to other localities, whereas a study of a New York City or another global outlier would beg the question of relevance for studying metropolitan areas more broadly. Second, Louisville is an increasingly prominent player in the landscape of data-driven urban governance, with Mayor Greg Fischer receiving national and international recognition for various policy initiatives aimed at making data, including data from social media platforms, a key driver in municipal policy development and implementation (Carroll 2013; Fischer 2012; Goldsmith 2013; Louisville Metro

Government 2012; Reno-Weber and Niblock 2013; Shelton et al 2015). Third and finally, Louisville is a city with intense social inequalities and a keen appreciation how they are manifest spatially. This is seen most clearly in the notion of the "9th Street Divide", which signifies the material inequalities and imaginative distance that separates the city's predominantly poor and African-American neighborhoods in the West End from more affluent and predominantly white areas throughout the rest of the city (Crutcher 2013).

As such, this study provides an opportunity to show how big data can be mobilized to produce alternative understandings of cities and urban processes. It is, however, important to acknowledge that our choice of case study is not accidental, and that the insights gleaned from our analysis rely on our own local knowledge and understandings of the city's social dynamics, taken from our experiences living in and conducting a variety of research projects in Louisville. Our choice to highlight this is, however, much more than simply an acknowledgement of our own situatedness and biases; it is also an explicit attempt to counter the notion that meaningful insights about cities can be gleaned simply by 'crunching the numbers'. Understanding urban socio-spatial processes requires more than massive amounts of data and clever software algorithms; it also necessitates grounded understandings of local history and culture, and the broader political-economic forces at play. Thus, our goal in this paper is to highlight the usefulness of combining the conceptual approaches of critical socio-spatial theory with new methodological approaches being utilized to understand big social media data.

II. Information Technologies and the Contemporary Urban Condition

Though the use of new sources of data and other new technologies are at the center of many contemporary urban policy initiatives, information technologies have long played a prominent role in the way that urban spaces are conceived, planned and enacted. This is especially true of mapping and geographic information technologies, whether in the form of hand-drawn maps or Google Maps mashups used to display data interactively on the web (Schein 1993; Soderstrom 1996). And while these technologies have evolved from early computer models and planning support systems towards more participatory and web-based approaches to GIS, the nascent 'smart cities' movement has begun to shift these technologies from desktop computers towards being embedded in the fabric of the city itself, allowing for a continuous collection and analysis of heterogeneous data streams meant to make urban systems operate more rationally and efficiently (Greenfield 2013; Kitchin 2014).

A. Urban Analysis in the Era of Web 2.0 and Big Data

One of the most powerful ways that information technology is shaping urban life in the 21st century is through the production of digital content – text, photos, videos, etc. – tied to particular locations on the earth's surface. While the act of creating a geotagged tweet, posting a photo to Instagram, reviewing a restaurant on Yelp or 'checking in' to your favorite park on Foursquare may seem relatively mundane, these platforms and data sources are allowing for new ways of interacting with, and studying, cities (Arribas-Bel 2014). As both Goodchild (2007, 2009) and Graham (2010) have argued, these platforms of data production offer unprecedented possibilities for codifying local knowledge about otherwise neglected places and making it widely accessible, even opening up the possibility for non-positivist epistemologies of mapping (Elwood and Leszczynski 2013; Warf and Sui 2010). These platforms not only allow for such local knowledge to be transferred to or accessed from distant places, but they also allow citizens in close proximity to one another to interact in a place-specific way through digital networks (Hardey 2007).

While this data can be incredibly important for helping tourists navigate through unknown places using their smartphones and a combination of location-based applications, the significance of this data for the purposes of this paper is our capability of collecting, aggregating, mapping and analyzing this data to understand how these digital data shadows are intimately intermingled with offline, material geographies of everyday life. Geotagged social media data has been used to research topics ranging from linguistic and religious differences (Graham and Zook 2013; Shelton et al 2012; Wall and Kidnark 2012; Watkins 2012), to differences in the places frequented by locals and tourists in different cities (Fischer 2010; Poorthuis 2010). Others have used this kind of data to rethink how we conceptualize and define neighborhoods or other spaces of social affinity. For instance, Cranshaw et al's (2012) 'Livehoods' project used

Foursquare check-ins to re-draw neighborhood boundaries based on similarities in user mobility, while Stefanidis et al (2013) expand to the broader scale of the nation to understand the ‘polycentric communities’ formed by attention to and engagement with current events in far off places. Others have attempted to explain spatio-temporal variations in this content, with attention to processes such as the weekly movement of students leaving a college town in search of other entertainment options (Li and Shan 2013; see also Li et al 2013). Kelley (2013) uses a mixed method analysis to understand not only the influence of social inequality on shaping the data shadows within particular urban neighborhoods, but also the different perceptions and experiences of place embodied in such content.

Despite the range of issues this data can be used to address, it’s important to keep in mind that offline, material social processes, such as persistent social inequalities, continue to shape the data as we interact with it, never including everyone equally or in a representative fashion (Crutcher and Zook 2009; Graham et al 2014; Graham and Zook 2011). Even those who create such content might be marginalized through the voices of other contributors, or through the automatic sorting of software algorithms that judge a particular comment to be of lower value (Zook and Graham 2007). As such, it is important to recognize the limits of such technologies, as they approach only an incredibly shallow vision of democratization, if any (Haklay 2013). Nonetheless, these massive streams of real-time social data are being incorporated into automated systems collecting information on energy use, traffic congestion and any number of other urban processes, which are then used to make decisions about both the day-to-day operations and long-term planning of the city. From smartphone apps collecting data on potholes in order to prioritize areas in need of infrastructure maintenance in Boston to the monitoring of African-American teenagers' social media activity in Louisville, these kinds of applications are increasingly popular among local governments (Crawford 2013; Leonard 2014). That is, this data is not only interesting for its ability to shed light on relatively mundane geographic processes; it is also being used to directly shape the way we live in cities today.

B. Cities and Social Media Beyond the Geotag: Re-engaging Socio-spatial Theory

A key shortcoming to both scholarly and applied uses of this kind of data is, we believe, the failure to capture the broader range of socio-spatial processes that are embedded in the data. While the practice of geotagging only allows for these pieces of content to be tied to a single pair of latitude and longitude coordinates, this single point does little to reflect the variegated and polymorphous geographies of everyday life that this content represents (Crampton et al 2013). Though many of the aforementioned studies have attempted to mobilize more complex understandings of space and socio-spatial processes, we believe a more direct engagement with longstanding theoretical approaches to be a fruitful way to push forward research into the geographies of social media data.

As such, this paper specifically seeks to integrate the concepts and disposition of what might broadly been termed 'relational socio-spatial theory' into our analysis.

Taking off from Massey's original conceptions of a 'global sense of place' (1991), this work attempts to conceive of space as networked, fragmented and processural, rather than as a kind of fixed container with defined boundaries and characteristics. From reconceptualizations of globalization (Amin 2002) to a new focus on mobility as a fundamental, defining characteristic of contemporary life (Sheller and Urry 2006), a key tenet of this approach has been an inversion of Tobler's so-called 'first law of geography' – that all things are related, but near things are more related than far things. Instead, relational approaches suggest that "we cannot assume that local happenings or geographies are ontologically separable from those 'out there'" (Amin 2002: 386). By focusing on the social relations that recursively produce space and are in turn influenced by it, rather than simply privileging proximity in absolute, Cartesian space, Amin argues that we can begin to see "a subtle folding together of the distant and the proximate" (2007: 103). As social processes are more and more spatially extensive, owing at least in part to the increasing prevalence of information and communication technologies, our spatial categories similarly need to evolve so as not to assume universal connections between social activities or processes and the locations on the earth's surface at which they occur. And while much of this work has been produced with specific reference to cities and the urban, it has also been used with respect to the broader spatial scale of the region (Allen and Cochrane 2007; Amin 2004), as well as to the sub-urban scale, as evidenced in Massey's original focus on her own neighborhood of Kilburn in London.

Though much of this ‘relational turn’ sees itself as counter-posed to the conventions of Marxist political economy approaches (cf. ongoing debates in Amin and Thrift 2002; Brenner et al 2011; MacLeod and Jones 2007; McFarlane 2011), an attempt to return to the work of Henri Lefebvre and his understanding of ‘planetary urbanization’ has provoked a similar tendency to view the urban ‘without an outside’, as unevenly stretched across the space of globe through networks and flows that have come to define global capitalism (Brenner 2013; Brenner and Schmid 2014). Despite this lingering conflict between ostensibly opposed epistemological and ontological positions, work by Jessop et al (2008) and McCann and Ward (2010) has attempted to highlight the compatibility of these relational and territorial approaches, and the potential for combining these insights in geographical analysis. Indeed, with respect to the analysis of social media data, Shelton et al (2014) have previously mobilized Jessop et al’s Territory-Place-Scale-Networks framework as a means by which to account for the multidimensionality and polymorphous nature of socio-spatial relations. Ultimately, this kind of relational approach is useful in that it does not attempt to impose arbitrary limits on one’s analysis. By broadening the scope of one’s geographical imaginary to processes that span localities and to places quite distant in absolute terms, one can gain a greater appreciation for and understanding of the social and spatial context in which contemporary social activities are situated.

III. Methodology and Data Collection

In order to operationalize the orientation of relational socio-spatial theory with respect to social media data, we draw on methods from qualitative and critical GIScience, taking advantage of the power of both qualitative and quantitative analysis of this data, and the utility of mapping and geovisualization for communicating such analyses. In addition to these connections, this work also draws on work which seeks to understand urban inequalities not simply through official statistics related to race/ethnicity, income, education and other indicators geolocated to one's official residence and aggregated to more-or-less arbitrary spatial units. Ahas et al (2010) and Silm and Ahas (2014) have previously demonstrated the alternative possibilities for understandings of inequality based not on these more-or-less static statistics, but on measuring people's movements through cities and identifying the places that they actually inhabit on an everyday basis, and then understanding the deeper meanings behind such patterns (see also Kwan 2013; Wong and Shaw 2011).

In order to combine the attention to individuals as active, conscious producers of their own everyday lives, while also understanding the socio-spatial context that structures these activities and interactions (Hägerstrand 1970; Pred 1984), we developed a method that allows us to move between individual-level analysis of Twitter users and a neighborhood or area-based analysis that situates these users in the spatial context of sub-urban areas within the city of Louisville. To begin, we collected all geotagged tweets with exact latitude and longitude coordinates from within Louisville, Kentucky from late June 2012 to early July 2014, yielding a total of 5.7 million tweets (see Figure 4.2a, which represents a 1% random sample of these tweets for the year 2013). While this kind of map is a useful starting point, such visualizations of 'dots on a map' do little to overcome the understanding of social media data being defined simply by its latitude and longitude coordinates. This is a trend recently pushed to its logical conclusion by what cartographer Kenneth Field has called 'animated ectoplasm maps' of geotagged tweets generated with easy-to-use online mapping tools (Field 2014), but which allow for very little substantive understanding of geographical processes. Rather than simply describing the presence or absence of data points, our goal is to explore how each individual data point might be associated with other data points through spatial or social proximity, as well as situated within both the immediate and broader spatial context in which it was created.

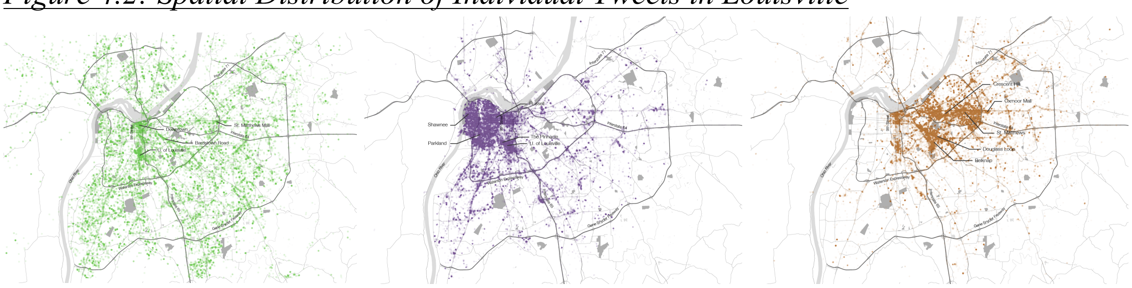
In order to do this, we refined our dataset by drawing boundaries around the West End and the East End, using the same classifications as a popular local magazine article comparing the two areas (Crutcher 2013; see Figure 4.1), and selecting only those tweets within the two areas, yielding approximately 450,000 tweets in the West End and 1.1 million tweets in the East End. But simply noting that there is a relative dearth of tweeting in the West End only serves to affirm more general notions of a persistent digital divide between such predominantly poor and African-American neighborhoods and those predominantly affluent and white areas of the East End. So, in order to understand the everyday geographies of individuals living in, or at least spending considerable amounts of time in, these neighborhoods, we devised a step-by-step approach in which users are classified as 'belonging' to one of the two areas.

Figure 4.1: West End and East End Boundaries Used for Data Collection

To meet this definition, users were required to satisfy two criteria. First, we identified those users expected to have created at least 40 geotagged tweets from within one of the two areas, based on a 10 percent random sample of tweets from within each neighborhood. This threshold is chosen to select relatively active Twitter users and to make sure 'belonging' is not based on only a few tweets that happen to be sent from within the neighborhood. Second, for users that satisfied the first criteria, we subsequently collected *all* of their tweets from within the city – both within and outside of their ‘home’ neighborhood – and further selected only those users with at least 50 percent of their geotagged tweeting activity originating within one of the two neighborhoods.

These criteria are meant to ensure the inclusion of only those users who predominantly and persistently tweet from one of these two areas, while excluding users with short-lived or minimally used accounts, as well as any one-time or infrequent visitors to one of the areas in question. The 50 percent threshold also ensures that no users are classified as being both West End *and* East End users. The end result is a dataset consisting of 703 users from the East End generating 274,338 geotagged tweets and 662 users from the West End with a total of 398,432 geotagged tweets (see Figures 4.2b and 4.2c). This dataset forms the foundation from which we are able to analyze aggregate patterns of socio-spatial mobility and segregation through a comparison of daily activity spaces of these two different groups.

Figure 4.2: Spatial Distribution of Individual Tweets in Louisville



In order to situate each data point within its broader spatial context and allow for comparisons between the selected neighborhoods, we aggregated each individual data point to a hexagonal grid composed of units roughly one kilometer in diameter (Scott

1985; Carr et al 1992). A key part of this aggregation process was adjusting the absolute count of tweets in each hexagonal area to control for the influence of 'power users', so that the resulting visualization was more reflective of the entire user base (cf. Poorthuis and Zook 2014). This decision reflects the social practices surrounding Twitter, in which a small group of users produce a disproportionally large amount of tweets; an issue that is compounded when examining specific subsets of tweets restricted by time or space. For instance, within the larger dataset, one user created 65 tweets from the area around 2nd and Market Streets in Louisville in one six hour period but never again tweeted from this area. Unadjusted, this activity would give equal weighting to each of these 65 tweets as to the tweets of individuals who travel regularly to this place, or individuals who only visit once but produce a much smaller amount of content. While this may be appropriate for some kinds of studies – particularly those focuses on individual behavior – our focus on the broader social and neighborhood context led to our approach to only include a maximum of five randomly selected tweets per user in any given hexagon. In other words, an adjusted tweet count of 50 for a given hexagon signifies that there were *at least* ten different users tweeting from that area. The sample size of 5 was chosen to provide some indication of the strength of a user's relation to that specific place while not diverging too far from the median value of one tweet per user per hexagon. This resulted in a total of 50,948 adjusted tweets from East End users, and 50,451 adjusted tweets from West End users. These adjusted counts, still aggregated to the hexagonal bins, serve as the basis for the calculation of the odds ratio measure seen in Figure 4.3 below.

IV. Visualizing Urban Socio-spatial Inequalities Using Social Media Data

Louisville, Kentucky is, like nearly all American cities, still marked by the legacies of racial segregation. From discriminatory housing policies of the mid-20th century that forced black residents into the city's West End neighborhoods, to ongoing struggles over the busing of school children, segregation and attempts to fight it have been at the center of Louisville's historic development[10] (Blum 2006; Cummings and Price 1997; Poe 2013). Ultimately, these structural forces produced a city increasingly divided along lines of race and class (Louisville Metro Human Relations Commission 2014), understood in the collective geographical imagination through the lens of the so-called "9th Street Divide", a colloquialism referring to the traditional boundary between the city's predominantly poor and black West End neighborhoods and the central business district.

Since the election of Mayor Greg Fischer in 2010, the West End has received an increasing amount of attention from the municipal government, with one Metro councilman from the almost entirely white and affluent far East End suburbs even declaring that "the 300,000 people east of Bowman Field do not exist in terms of what the mayor thinks about" (Bailey 2014). From issues of vacant and abandoned property

[10] It is also worth noting that these struggles in Louisville have played a key role in the broader national experience with racial segregation. Both the 1917 U.S. Supreme Court decision in *Buchanan v. Warley*, as well as the 2008 decision in *Parents Involved in Community Schools v. Seattle School District No. 1*, were centered on practices of state-sponsored racial segregation in Louisville.

to air quality, the city has begun investing more time and money in addressing the range of challenges faced by these neighborhoods, in some ways spurred by a recent cover story in *Louisville Magazine* documenting the staggering inequalities between the West End and the rest of the city (Crutcher 2013). A key element of the collective social imagination of Louisvillians, as noted in both the *Louisville Magazine* story and in other venues, is the notion that the West End is fundamentally separate and apart from the rest of the city, with the aforementioned problems being isolated within these areas; the rest of the city does not have to deal with the ill effects of these issues, nor did they help to create them. The conceptualization of an isolated West End is reinforced by discussions of intra-urban mobility more generally. For example, in a presentation to the Louisville African American Initiative's West Louisville Economic Development and Housing Summit on July 11, 2014, Fischer argued that mobility is one of the most persistent problems in Louisville, leading people of all ages to remain segregated; people from the East End tend to stay in the East End, people from the West End stay in the West End, people from the South End stay in the South End.

Because such understandings of the West End and its relationship to the rest of the city are clearly at conceptual odds with the relational understanding of space we use in this paper, we use the rest of the paper to empirically document that the conventional wisdom around socio-spatial segregation in Louisville is overly simplistic. While we do not discount the importance of both public policy and individual actions in producing the uneven landscapes of segregation in the city today, we demonstrate that the West

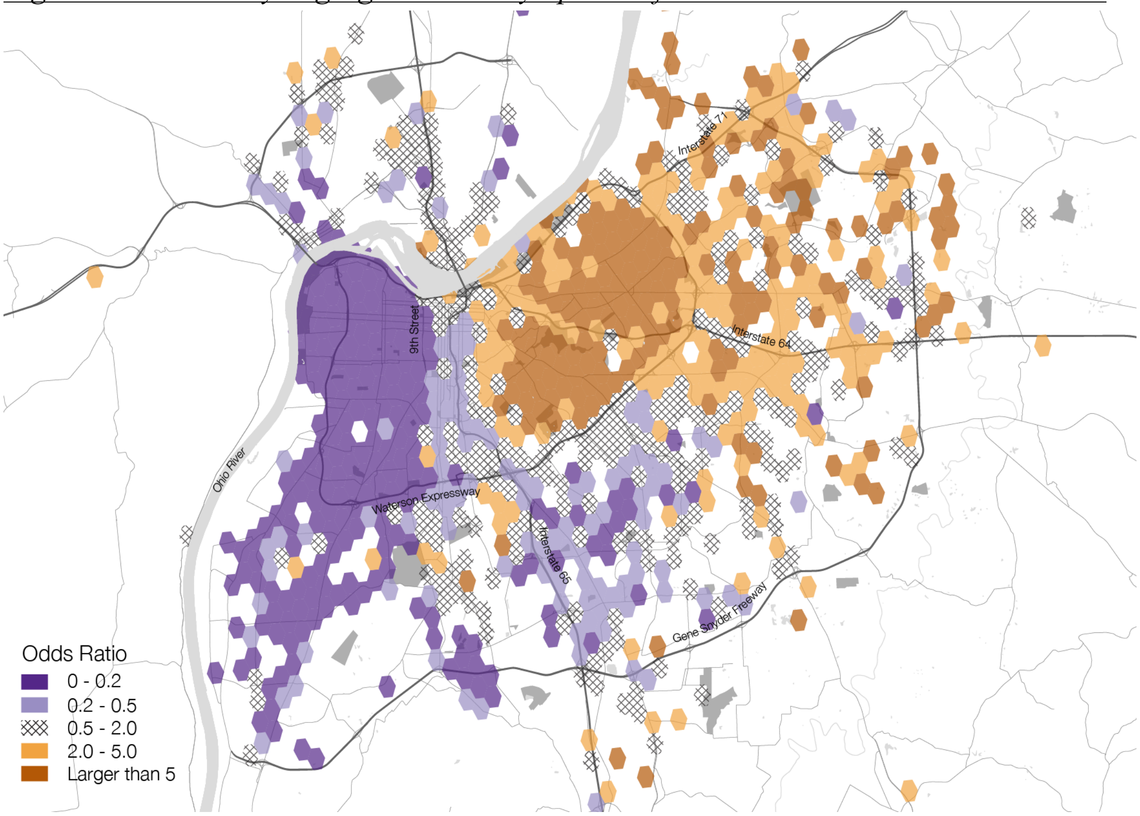
End, through its relations to the rest of the city and the everyday mobilities of the people who live there, is not defined by segregation alone, nor is this process of segregation total. It is rather a complex, partial and selective process that affects some people and places more intensely than others.

A. Everyday Mobilities and Activity Spaces

The first step in our analysis is analyzing the relative patterns of mobility of each of our two groups: West End and East End Twitter users. To do so, we calculated the odds ratio of each of the hexagonal cells spanning the greater Louisville area, using the adjusted tweet count, in order to show the relative presence or absence of each of the two groups throughout the city (see Figure 4.3). The odds ratio indicates the prevalence of one group relative to the other in a particular place, while adjusting for the overall amount of tweeting from each of the two groups in that place. Values approaching 1 indicate that there is relative parity, while values approaching 0 showing a much greater prevalence of West End tweeting and values greater than 1 showing a greater prevalence of East End tweeting.

As expected, the areas used as the bounding boxes for classification of the two neighborhoods both show a general dominance of those users. But more useful insights can be gained by looking at the extent to which locations outside of these areas demonstrate high levels of either heterogeneity or homogeneity in tweeting activity and, thus, physical presence of West End and East End users. At first glance, the 9th Street Divide is clearly evident as the boundary at which a significant dominance of West End users gives way to greater levels of heterogeneity in the downtown area. But since the idea of the 9th Street Divide is generally used to denote the ‘failings’, so to speak, of the West End, including the segregation and constrained mobility of its residents, it is important to note that it is actually East End users who are more spatially segregated and confined in their everyday activity spaces, which will be discussed more below.

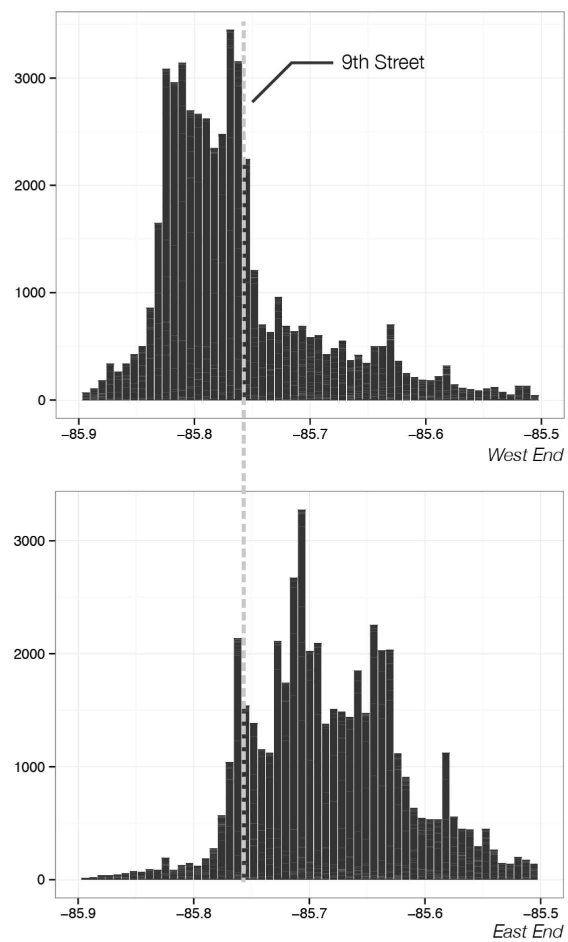
Figure 4.3: Unevenly Segregated Activity Spaces of West End and East End Residents



Given the large number of places with a relative prevalence of West End tweeting that are quite distant from the West End, it is evident that West End Twitter users are relatively mobile within the Metro area, traveling across the Ohio River to southern Indiana, or to neighborhoods like Hikes Point, on the periphery of the East End as we have defined it here. Indeed, only focusing on hexagons with odds ratios less than 0.5 as indicators of West End users' mobility actually tends to understate the prevalence of West End users in other parts of the city (compare areas in purple to concentrations seen in Figure 4.2b). Traces of this spatial mobility are also evident in isolated areas around major travel routes, such as I-64 or I-264, in the East End, where there is a lower relative amount of East End tweeting than in the surrounding residential areas. This initial look is interesting insofar as it offers a counterpoint to the conventional narrative that class and racial segregation in housing is mirrored by limited mobility by these groups (cf. Cass et al 2005), especially in a city without significant alternative transportation options such as Louisville, though of course such mobility by West End residents is likely driven by the necessity of seeking out employment and consumption activities outside of their home neighborhoods.

In contrast, residents of the predominantly white and relatively affluent East End exhibit constrained activity spaces in that there are many parts of the city, *especially* the West End, to which they do not travel. For example, 45% of the adjusted tweets from West End users were east of 9th Street, while only 7% of tweets from East End users were west of 9th Street (see Figure 4.4). This contrast is all the more stark when accounting for the fact that most of the South End neighborhoods, which demonstrate the most significant connection with the West End, are also west of the 9th Street boundary. The lack of presence of East End Twitter users is particularly intriguing given that the West End has become an increasingly important object of attention for policy elites (predominately from the East End) in the city. From a film and discussion series on urban renewal put on by a local historical society to a workshop about the rehabilitation of vacant properties in the West End, there are numerous examples of the public discourse around the West End that are dominated by individuals or groups with little or no direct connection to these neighborhoods. Indeed, it was the influence (or lack thereof) of outsiders that was seen a primary determinant in Metro Councilwoman Attica Scott's failed reelection bid in the 2014 primary elections.

Figure 4.4: Distribution of Tweets to the West and East of 9^{th} Street



Despite this broader interest in the current and future state of the West End, it is evident that many East End Twitter users rarely come into actual contact with the West End, seemingly avoiding these neighborhoods almost in their entirety. This isn't, of course, particularly surprising. As one local journalist summarized the relationship between the West End and East End:

> "Any Louisvillian who has lived here for more than a few years knows, almost instinctively, the boundary line between west Louisville and the rest of Louisville: Ninth Street. Most white Louisvillians know it because they've heard some variation of the warning, 'Don't go west of Ninth Street.'…Although the notion that west Louisville is a dangerous and even foreign place is embedded in the mental map that many of us – even the most bleeding-heartedly liberal and racially tolerant, if we will admit it – carry around in our heads, it is rarely talked about in public" (Crutcher 2013: 25)

Indeed, this is even reflected in more popular expressions, such as the closing line of one local's take on a popular meme, entitled "Shit Louisville People Don't Say", in which the white male narrator says ironically from the seat of his car, "Hey, do you want to go down to the West End?"[11].

Thus, despite the salience of the 9th Street Divide metaphor, and its broader importance in drawing attention to socio-spatial inequality in Louisville, it clearly belies the complexity of how these inequalities are manifest in the everyday activity spaces of Louisvillians. When local real estate developer and civic leader Gill Holland called 9th Street "the Berlin Wall of our community", he likely meant to reinforce the notion that it is a rigid boundary. But like the Berlin Wall, the 9th Street Divide is not insurmountable, though movement across such divides were and are largely

[11] Video available from: http://youtu.be/7Ru1qNDSP-o.

unidirectional, encouraging some to cross, while others remain prohibited or discouraged from such movements based on their social positions and broader structural forces. West End residents regularly travel outside of their home neighborhoods towards educational, employment and consumption opportunities further east and south, while many fewer East End residents move to their west when performing similar activities. These dynamics suggest the need for a more nuanced understanding of the 9th Street Divide, not as a border with fixed, defined and unchangeable characteristics that physically divides the city, but as a kind of spatial imaginary that isn't entirely mirrored in people's material spatial practices.

B. Exploring the Fluidity of Neighborhood Boundaries

Building upon our analysis of the activity spaces of West and East End Twitter users, we now turn towards a broader reinterpretation of the neighborhoods and their boundaries based on the everyday mobilities of residents. Figure 4.5 provides a simple visualization of those hexagons with 50 or more adjusted tweets for each user group, so as to provide a more expansive and spatially extensive definition of these neighborhoods than is possible with the use of conventional census tracts or area units. Our redrawing of these boundaries points toward the fluidity and porousness of the neighborhoods; while the West End and East End remain spatially distant enclaves in some respects, they also overlap at key points, such as the downtown and waterfront area, as well as suburban malls.

Figure 4.5: Redefining the Boundaries of the West End and the East End

Our redefinition of the East End is rather subtle, largely a result of the fact that many more tweets from East End users fall within our original boundaries than is the case for West End users (compare Figures 1 and 2c). But East End users, by and large, tend to gravitate eastwards, towards the outer suburban areas of Hurstbourne and Middletown, and commercial areas like the upscale mall The Summit. In addition, this redefined spatiality of the East End encompass much of the city's downtown area as well as the traditionally working-class white neighborhoods of Germantown and Schnitzelburg, suggestive of recent urban redevelopment seeking to draw people back to the city's downtown. The only point at which our expansive East End boundary approaches the conventionally defined West End is in the areas surrounding the Churchill Downs racetrack and the University of Louisville (see Figure 4.7 below).

The redrawing of the West End is, however, much more significant in scope. While the census tract definition is bounded by 9th Street to the east, Algonquin Parkway to the south, and the Ohio River to the north and west, the everyday mobilities of West End Twitter users extend throughout the city. From the entire downtown area and the University of Louisville campus, to major transportation and commercial corridors throughout the East End, our redefinition of the West End demonstrates the incredibly partial story told through tropes of the West End as being somehow isolated and apart from the rest of the city.

For example, the activity spaces of the West End highlight a strong connection to the South End neighborhoods, a predominantly white and working class community sometimes characterized as the 'redneck' part of town in same way the West End is pejoratively labeled as 'the ghetto'[12]. For a city with such stark racial differences, it seems counterintuitive that the South End would demonstrate a more significant connection to the West End through such everyday mobilities than with the similarly white and suburban areas of the East End. But, we would argue, this points towards the importance of combining such analyses of big data with situated, place-based knowledges that allow for an explanation of such anomalies, even if such an explanation isn't definitive. In this case, the connection traces back to the movement of white working class families from the West End to the South End following the Great Flood of 1937, and the resulting influx of black families into these neighborhoods after WWII, itself a key moment of racial conflict in the city (Welch 2013). Extending the metaphor of this kind of unstructured information as 'data exhaust', we see the extension of the West End southward in Figure 5a as the digital contrails of white flight, which continue to shape patterns of mobility within the city.

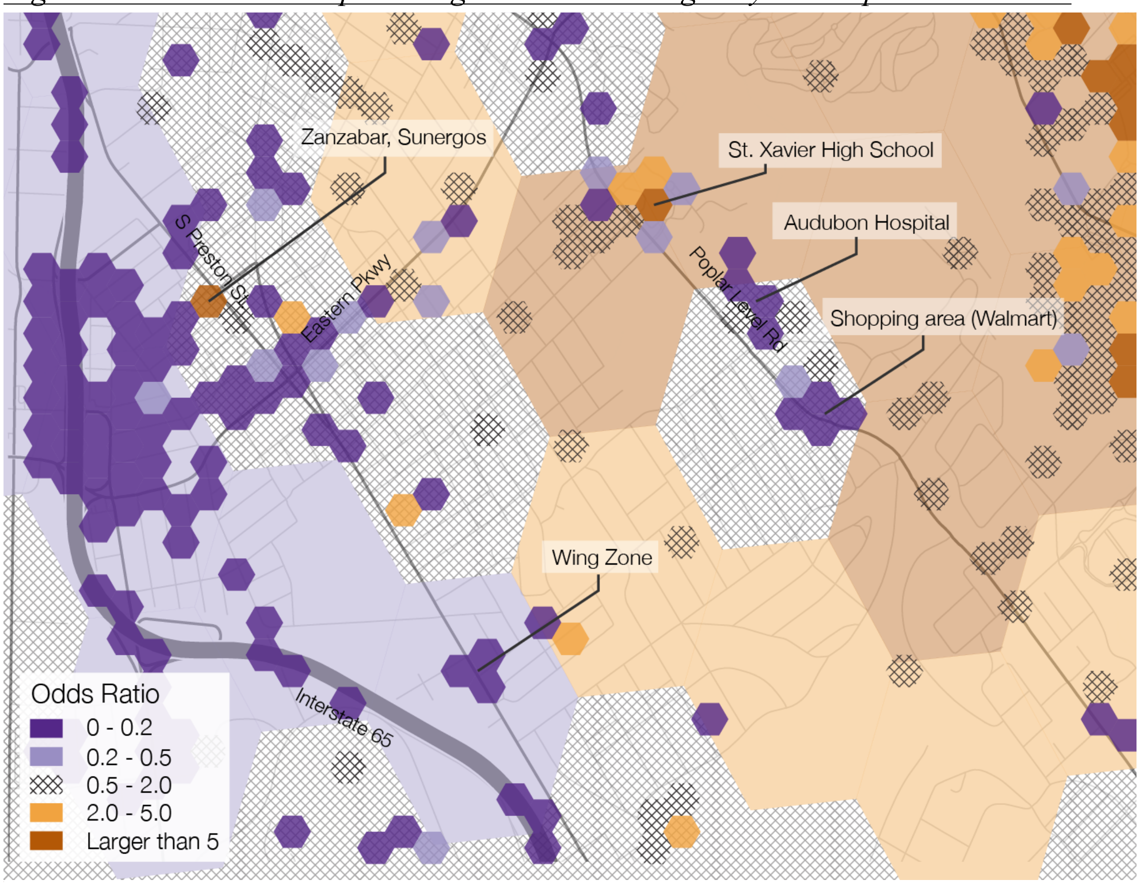
[12] For example, see the map of Louisville from the blog Judgmental Maps, available from: http://judgmentalmaps.com/post/83423132066/louisville.

C. Scale Dependent Understandings in Space and Time

Thusfar, we have largely highlighted those areas that represent differences in the spatial patterns and practices of geotagged tweeting by West End and East End users. We now turn to exploring in more depth those places that represent greater social heterogeneity in the city, as defined by greater parity in the levels of West End and East End tweeting in a given locality. But just as conventional understandings of the 9th Street Divide belie the complexity of relations between West End residents and the rest of the city, so too can our earlier analysis and methods of classification disguise what are still highly fractured social spaces as areas of heterogeneity and social mixing. While we can identify a number of relatively heterogeneous areas throughout the city (see areas in Figure 4.3 represented by the hatched pattern for location quotient values between 0.5 and 2.0), this understanding is largely shaped by the methods we have utilized. By mobilizing alternative scalar framings (cf. Feick and Robertson 2014), we can see that while West End and East End users are often incredibly spatially proximate to one another, they are rarely using the same physical spaces, and are even more unlikely to be co-present in the same places at the same times.

Using a finer scale analysis of one area of the city defined by relative heterogeneity, Figure 4.6 demonstrates the significant 'splintering' effect made visible by this kind of social media data and a multi-scalar analysis (Graham and Marvin 2001). In our initial analysis in Figure 4.3, this area is characterized by varying levels of West End and East End tweeting, with some hexagons displaying relative heterogeneity, while others demonstrate a strong preference for one group over another. This is perhaps best illustrated in the block of businesses along South Preston Street (shown towards the upper left corner of Figure 4.6), including Zanzabar, The New Vintage and Sunergos Coffee Shop which are all dominated by East End tweeting, consistent with their business models which cater to a younger, 'hipster' demographic in the neighborhood. In contrast, tweets from West End users in the area are much less concentrated, and tend to be located off of main thoroughfares in residential areas.

Figure 4.6: Multi-scalar Splintering near Preston Highway and Poplar Level Road



Similarly, the area surrounding Audubon Hospital and a small shopping area (seen in the right side of Figure 4.6) shows that at this scale, there are a couple of smaller areas of heterogeneity, surrounded by significant concentrations of West End tweeting. In addition, at this scale one can clearly see how the scalar shift influences measures of heterogeneity, as Audubon Hospital sits near the boundary of two hexagonal cells and is proximate to a largely wealthy and white, all-boys Catholic high school, with a dense concentration of East End tweeting, but which also straddles two hexagonal areas.

Such multi-scalar patterns of splintering urbanism are also evident at different *temporal* scales. Like the above analysis, some places that appear to be heterogeneous are actually marked by different uses across time, which can also impact our understanding of what's actually happening on the ground in these locales. One such place is Churchill Downs, the historic horseracing track located just to the southeast of the traditional West End boundary. In Figure 4.3, Churchill Downs stands out as the lone cluster of East End tweeting west of Interstate 65 and inside of the Interstate 264 loop, and is surrounded by hexagons displaying varying levels of West End tweeting. But since much of the East End user presence in this area is related to horseracing, itself a seasonal activity, this locale offers the potential to demonstrate the temporal splintering in activity spaces across our two user groups. While it is broadly evident that much of the East End tweets in the area are in and around the track, with West End tweets more dispersed, we subdivided tweets in this area based on those occurring in and around racing season (April to June and September to November), and those in other months when the track is not active.

Figure 4.7: Temporal Differences in Tweeting Near the Churchill Downs Race Track

Figure 4.7 uses a finer spatial scale similar to that seen above in Figure 6 to show the relative amount of tweeting between West End and East End users during racing season (Figure 4.7a) and outside of racing season (Figure 4.7b). While there are many more West End tweets (n=4,191) than East End tweets (n=1,355) in this area, both user groups tend to tweet more from these areas during racing season, and at nearly equivalent rates. It is evident, however, that the places these users tweet from during the two time periods are significantly different; East End users tweet largely from Churchill Downs and the nearby Papa John's Cardinal Stadium during the months of racing season, while these concentrations diminish in non-racing months. West End users tend not to tweet from these places in any significant amounts, with tweets distributed throughout the surrounding neighborhood. We can thus see the dynamism of the area around Churchill Downs through time: most often this area is a spatial extension of the

West End, though it temporarily transforms into a space of (relatively) elite consumption and a site of global attention, despite still being surrounded by a neighborhood from which many visitors have little connection whatsoever.

D. Contextualizing Data Practices Through Qualitative Analysis

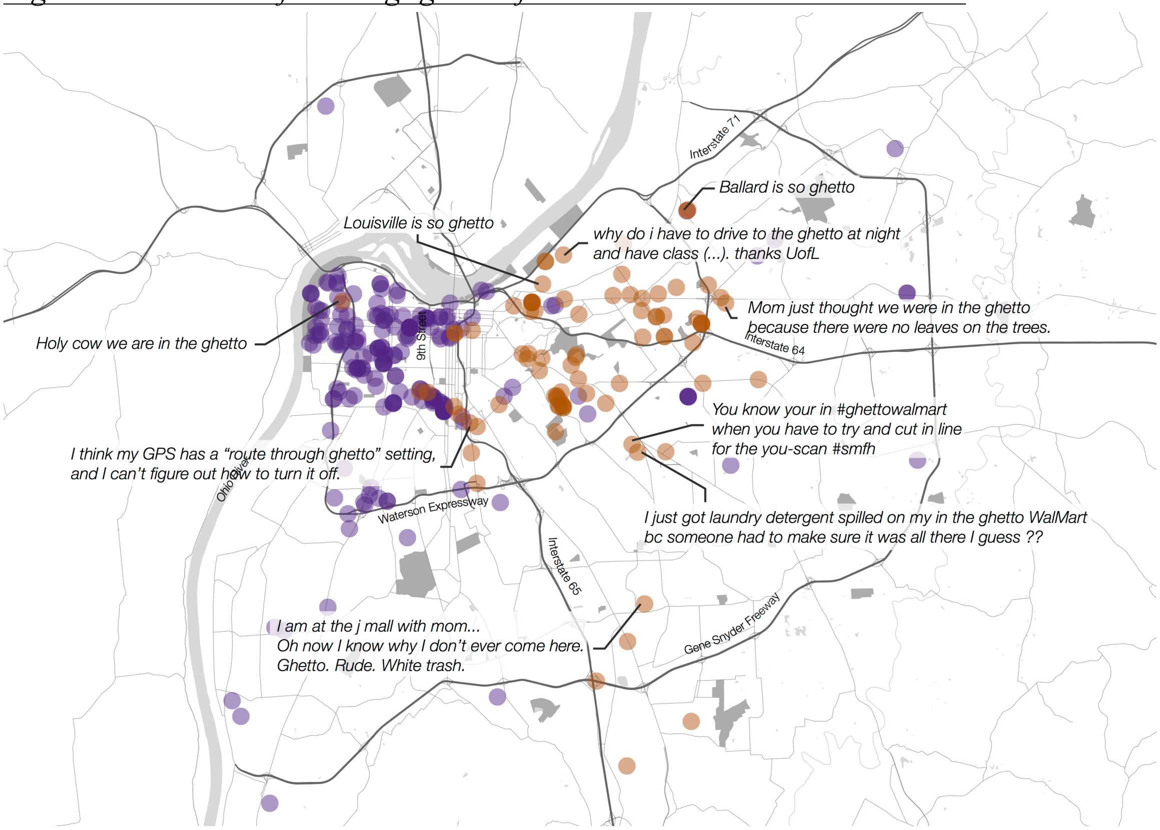
Up to this point, much of our analysis has followed in the tradition of research using mobile phone records or GPS trackers to study everyday mobilities and segregation. Apart from the fact that the use of geotagged social media data is considerably more cost-effective and less invasive for longer-term studies of mobility, one of the biggest advantages of using the massive databases of social media data available to us, however, is that it allows for greater attention to social context, rather than simply providing a record of presence in a given location at a given time. With geotagged Twitter data, we have some insight into what a given person was thinking or talking about, how they describe themselves, and with whom they were communicating, among other things. The qualitative content embedded in each individual tweet thus provides another means by which we can understand how Louisvillians move through, inhabit and experience the city.

To further target our qualitative analysis of tweets, we collated all tweets our user groups that fell within the boundaries of hexagons we determined to be 'highly heterogeneous' – defined by a location quotient between 0.8 and 1.25 and a total of 100 tweets from West End and East End users combined – resulting in totals of 1,812 and 1,690 tweets, respectively. Hoping to further explore experiences of heterogeneity in places around the city, we conducted an inductive and iterative analysis of how users were engaging with these places, while keeping in mind that there is no necessary relationship between the location in which a user produces a tweet and the content of the tweet itself.

The first pattern that emerged from this analysis was the significant difference in the ways users' tweets were connected to Foursquare check-ins, an often associated geosocial media platform focused on location-based social networking. In this sample of tweets in highly heterogeneous areas, 10.7% of all East End user tweets (n=193) were associated with Foursquare check-ins, from a total of 44 separate users. On the other hand, there were only four Foursquare check-ins from four separate West End users, representing just 0.2% of all West End user tweets in these areas. This disparity, combined with the socio-spatial splintering noted earlier in Figures 6 and 7, suggests that the socio-spatial practices and imaginaries of these two user groups are fairly disparate, even when they are operating in physically proximate areas.

While somewhat speculative, we would suggest that the relatively high Foursquare activity by East End users is consistent with theories of 'conspicuous mobility', or "[the] re-figuring of everyday mobility as a consumptive activity" (Wilson 2012: 1271) based on the sharing of (particular kinds of) locational information. In this case, the check-ins by East End users highlight their presence at 'hip' places – coffeeshops, bakeries, restaurants, bars and music or other entertainment venues – and serve to make these mobilities and consumption practices more known to others. In contrast, West End users active in these heterogeneous areas demonstrate little of this kind of effort towards locational visibility and instead were more likely to engage in distanciated social interactions with individuals who are (seemingly) not co-present at that particular location, providing further weight to our understanding of the West End as being spatially diffuse and more appropriately defined by the density of social connections that stretch across urban space.

Figure 4.8: Tweets referencing 'ghetto' from West End and East End users



Another pattern we were able to identify was the relatively frequent references to 'the ghetto' and other associated terms amongst both East End and West End users. We chose to return to our original sample of East End and West End user tweets and map the spatial distribution of references to ‘ghetto’ in our entire corpus of tweets, which yielded 197 tweets from West End users and 87 from East End users. Interestingly enough, the spatial signature of the ‘ghetto’ in geotagged tweets amongst the two groups roughly mirrors the broader spatial patterns of these two users groups. There are very few East End tweets west of 9th Street, and while West End tweets are generally more concentrated within the conventional boundaries of the West End, there are also plenty of tweets outside of this area as well.

Figure 4.8 shows the distribution of these tweets, with West End tweets in purple and East End tweets in orange, with the text of selected tweets from East End users referencing the ‘ghetto’ also included. Looking closer to this content, we can see that only one East End user tweeted about the ghetto from the West End, saying “Holy cow we are in the ghetto”, while another user at Ballard High School in the East End declared “Ballard is so ghetto”. Especially in the case of these selected tweets, there is a level of cognitive dissonance at play, with many of these tweets from East End users being produced in predominantly white and affluent, and mostly suburban, areas, far distanciated both socially and spatially from anything that might resemble ‘the ghetto’. Ultimately, these kinds of incongruencies demonstrate the more complex relationship between urban spatial imaginaries and the everyday activity spaces of individuals and collectives as demonstrated through geotagged social media data.

V. Conclusion

In this paper, we have developed a conceptual and methodological approach to the study of geotagged social media data that responds to earlier calls to go 'beyond the geotag'. Rather than simply plotting the locations of individual tweets on a map, our approach combines relational socio-spatial theory with a variety of methods drawn from critical GIScience in order to place individual data points in relation to one another and to their broader social and spatial contexts through a more deliberate process of data collection and analysis.

In our case study, we focused on issues of intra-neighborhood segregation, mobility and inequality in Louisville, Kentucky, highlighting the fundamentally fluid, networked and relational nature of places in the city, as well as the dynamism of how people live in and occupy these places. Our analysis provides a strong counter-argument to the pervasive socio-spatial imaginary within the city of a '9th Street Divide' that tends to isolate and pathologize the West End and its residents. But by understanding how people from different parts of the city actually move through and experience the city differently, we are able to demonstrate the contours of a more complicated set of socio-spatial mobilities that define the city and its neighborhoods through their extralocal relationships to other people and places. While this work has focused in particular on rethinking the socio-spatial imaginaries connected to particular classed and racialized neighborhoods in Louisville, future research with this kind of data could just as well focus more explicitly on issues of age, gender, sexuality or other identities as inferred from user profiles. And though classifying individuals based on these demographics is challenging and beyond the scope of this paper, it represents an opportunity to build upon our understandings of how this data can reveal a more complex and nuanced set of socio-spatial relations than is typically assumed.

Ultimately, we wish to reiterate that we are not arguing that geotagged social media data is an unequivocal improvement on, or replacement for, other forms of social and spatial data, especially when analyzing questions of inequality. But rather than reinscribing these inequalities through the use of such datasets, we would argue that our analysis has shown that this kind of social media data represents a potentially rich source from which to construct empirically-grounded counter narratives of these inequalities and popular socio-spatial imaginaries thereof, which in turn can allow for alternative conceptualizations of, and interventions into, urban socio-spatial relations and processes.

VI. References

Ahas, Rein, Siiri Silm, Olle Järv, Erki Saluveer, and Margus Tiru. 2010. "Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones." *Journal of Urban Technology* 17 (1): 3–27.

Allen, John, and Allan Cochrane. 2007. "Beyond the Territorial Fix: Regional Assemblages, Politics and Power." *Regional Studies* 41 (9): 1161–1175.

Amin, Ash. 2002. "Spatialities of Globalisation." *Environment and Planning A* 34 (3): 385–399.

Amin, Ash. 2004. "Regions Unbound: Towards a New Politics of Place." *Geografiska Annaler: Series B, Human Geography* 86 (1): 33–44.

Amin, Ash. 2007. "Re-thinking the Urban Social." *City* 11 (1): 100–114.

Amin, Ash, and Nigel Thrift. 2002. *Cities: Reimagining the Urban*. Polity.

Anderson, Chris. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired Magazine* 15(7).

Arribas-Bel, Daniel. 2014. “Accidental, Open and Everywhere: Emerging Data Sources for the Understanding of Cities.” *Applied Geography* 49: 45–53.

Bailey, Phillip. 2014. “Does Mayor Greg Fischer’s Proposed Capital Budget Stiff Louisville’s East End?”. *WFPL 89.3fm.* 28 May. Available from: http://wfpl.org/post/does-mayor-greg-fischers-proposed-capital-budget-stiff-louisvilles-east-end.

Barnes, Trevor J. 2009. “‘Not Only...But Also’: Quantitative and Critical Geography.” *The Professional Geographer* 61 (3): 292-300.

Barnes, Trevor J. 2013. “Big Data, Little History.” *Dialogues in Human Geography* 3 (3): 297–302.

Batty, Michael. 2012. “Smart Cities, Big Data.” *Environment and Planning B: Planning and Design* 39 (2): 191–193.

Bettencourt, Luis, and Geoffrey West. 2010. “A Unified Theory of Urban Living.” *Nature* 467 (7318): 912–913.

Blum, Sarah Hardin. 2006. “Race, Housing, and the Making of Twentieth-Century Louisville, Kentucky”. Ph.D. Dissertation, University of Kentucky Department of History.

boyd, danah, and Kate Crawford. 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” *Information, Communication & Society* 15 (5): 662–679.

Brenner, Neil. 2013. “Theses on Urbanization.” *Public Culture* 25 (1): 85–114.

Brenner, Neil, David J. Madden, and David Wachsmuth. 2011. “Assemblage Urbanism and the Challenges of Critical Urban Theory.” *City* 15 (2): 225–240.

Brenner, Neil, and Christian Schmid. 2014. “The ‘Urban Age’ in Question.” *International Journal of Urban and Regional Research* 38 (3): 731–755.

Carr, Daniel B., Anthony R. Olsen, and Denis White. 1992. “Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data.” *Cartography and Geographic Information Systems* 19 (4): 228–236.

Carroll, James R. 2013. “At White House, Mayor Greg Fischer details how data improves lives”. *Louisville Courier-Journal*. 23 July.

Cass, Noel, Elizabeth Shove, and John Urry. 2005. “Social Exclusion, Mobility and Access.” *The Sociological Review* 53 (3): 539–55.

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. “Beyond the Geotag: Situating

‘Big Data’ and Leveraging the Potential of the Geoweb.” *Cartography and Geographic Information Science* 40 (2): 130–139.

Cranshaw, Justin, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. “The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City.” In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, June, pp. 58–65.

Crawford, Kate. 2013. “The Hidden Biases in Big Data.” *Harvard Business Review*. 1 April. Available from: http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html.

Crutcher, Dan. 2013. “A Tale of Two Cities”. *Louisville Magazine* (March): 25-29.

Crutcher, Michael, and Matthew Zook. 2009. “Placemarks and Waterlines: Racialized Cyberscapes in Post-Katrina Google Earth.” *Geoforum* 40 (4): 523–534.

Cummings, Scott, and Michael Price. 1997. “Race Relations and Public Policy in Louisville: Historical Development of an Urban Underclass.” *Journal of Black Studies* 27 (5): 615–649.

Elwood, Sarah, and Agnieszka Leszczynski. 2013. “New Spatial Media, New Knowledge Politics.” *Transactions of the Institute of British Geographers* 38 (4): 544–559.

Fairfield, John D. 1994. “The Scientific Management of Urban Space: Professional City Planning and the Legacy of Progressive Reform.” *Journal of Urban History* 20 (2): 179–204.

Feick, Rob, and Colin Robertson. Forthcoming. “A Multi-Scale Approach to Exploring Urban Places in Geotagged Photographs.” *Computers, Environment and Urban Systems*.

Field, Kenneth. 2014. “I'm wondering when people will realise the animated ectoplasm twitter maps don't actually show anything http://t.co/SJVYLyBn1F” [Tweet]. 17 June. Available from: https://twitter.com/kennethfield/status/478775510386741248.

Fischer, Eric. 2010. “Locals vs. Tourists”. Available from: https://www.flickr.com/photos/walkingsf/sets/72157624209158632/

Fischer, Greg. 2012. “Using ‘Big Data’ to Improve Public Health in Louisville.” *Citizen IBM*. 2 November. Available from: http://citizenibm.com/2012/11/using-big-data-to-improve-public-health-in-louisville.html

Ford, George. 1913. “The City Scientific.” *Engineering Record* 67 (May): 551–552.

Goldsmith, Stephen. 2013. “How Louisville, Ky., Is Using a ‘Stat’ Program to Transform the Culture of Government.” *Governing Magazine*. 19 June. Available from: http://www.governing.com/blogs/bfc/col-efficiency-louisvillelouiestat-performance-metrics-improvement-transform-government-culture.html.

Goodchild, Michael. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69 (4): 211–221.

Goodchild, Michael. 2009. "NeoGeography and the Nature of Geographic Expertise." *Journal of Location Based Services* 3 (2): 82–96.

Graham, Mark. 2010. "Neogeography and the Palimpsests of Place: Web 2.0 and the Construction of a Virtual Earth." *Tijdschrift Voor Economische En Sociale Geografie* 101 (4): 422–436.

Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. "Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty." *Annals of the Association of American Geographers* 104 (4): 746–764.

Graham, Mark, and Taylor Shelton. 2013. "Geography and the Future of Big Data, Big Data and the Future of Geography." *Dialogues in Human Geography* 3 (3): 255–261.

Graham, Mark, and Matthew Zook. 2011. "Visualizing Global Cyberscapes: Mapping User-Generated Placemarks." *Journal of Urban Technology* 18 (1): 115–132.

Graham, Mark, and Matthew Zook. 2013. "Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web." *Environment and Planning A* 45 (1): 77–99.

Graham, Stephen, and Simon Marvin. 2001. *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. Routledge.

Greenfield, Adam. 2013. *Against the Smart City*. Do Projects.

Hägerstrand, Torsten. 1970. "What About People in Regional Science?" *Papers in Regional Science* 24 (1): 7–24.

Haklay, Muki. 2013. "Neogeography and the Delusion of Democratisation." *Environment and Planning A* 45 (1): 55–69.

Hardey, Michael. 2007. "The City in the Age of Web 2.0: A New Synergistic Relationship between Place and People." *Information, Communication & Society* 10 (6): 867–884.

Jessop, Bob, Neil Brenner, and Martin Jones. 2008. "Theorizing Sociospatial Relations." *Environment and Planning D: Society and Space* 26 (3): 389–401.

Kelley, Matthew James. 2013. "The Emergent Urban Imaginaries of Geosocial Media." *GeoJournal* 78 (1): 181–203.

Kitchin, Rob. 2014. “The Real-Time City? Big Data and Smart Urbanism.” *GeoJournal* 79 (1): 1–14.

Kwan, Mei-Po. 2013. “Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility.” *Annals of the Association of American Geographers* 103 (5): 1078–1086.

LeGates, Richard, Nicholas J. Tate, and Richard Kingston. 2009. “Spatial Thinking and Scientific Urban Planning.” *Environment and Planning B: Planning and Design* 36 (5): 763–768.

Leonard, Connie. 2014. “Mayor calls for more social media monitoring”. *WAVE3 News*. 25 March. Available from: http://www.wave3.com/story/25073430/mayor-calls-for-more-social-media-monitoring.

Li, Linna, Michael F. Goodchild, and Bo Xu. 2013. “Spatial, Temporal, and Socioeconomic Patterns in the Use of Twitter and Flickr.” *Cartography and Geographic Information Science* 40 (2): 61–77.

Li, Yue, and Jie Shan. 2013. “Understanding the Spatio-Temporal Pattern of Tweets.” *Photogrammetric Engineering and Remote Sensing* 79 (9): 769–773.

Light, Jennifer S. 2003. *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*. Johns Hopkins University Press.

Louisville Metro Government. 2012. “Louisville Metro Government Named Top ‘Digital City’”. Press Release. 15 November. Available from: http://www.louisvilleky.gov/Technology/News/2012/11-15-12-digital_city.htm

Louisville Metro Human Relations Commission. 2014. *Making Louisville Home For Us All: A 20-Year Action Plan for Fair Housing*. Available from: http://www.metropolitanhousing.org/wp-content/uploads/member_docs/FairHousingReport_2013_15.pdf

MacLeod, Gordon, and Martin Jones. 2007. “Territorial, Scalar, Networked, Connected: In What Sense a ‘Regional World’?” *Regional Studies* 41 (9): 1177–1191.

Massey, Doreen. 1991. “A Global Sense of Place.” *Marxism Today* 35 (6): 24–29.

McCann, Eugene, and Kevin Ward. 2010. “Relationality/territoriality: Toward a Conceptualization of Cities in the World.” *Geoforum* 41 (2): 175–184.

McFarlane, Colin. 2011. “Assemblage and Critical Urbanism.” *City* 15 (2): 204–224.

Poe, Joshua. 2013. “A City Divided”. *LEO Weekly*. 2 January. Available from: http://leoweekly.com/news/city-divided.

Poorthuis, Ate. 2010. “Place in New York City and Amsterdam: Exploring the possibilities and limitations of Volunteered Geographic Information.” MSc. Thesis. University of Amsterdam.

Poorthuis, Ate, and Matthew Zook. 2014. "Artists and Bankers and Hipsters, Oh My! Mapping Tweets in the New York Metropolitan Region." *Cityscape* 16 (2): 169-172.

Pred, Allan. 1984. "Place as Historically Contingent Process: Structuration and the Time-Geography of Becoming Places." *Annals of the Association of American Geographers* 74 (2): 279–297.

Reno-Weber, Theresa, and Beth Niblock. 2013. "Beyond Transparency: Louisville's Strategic Use of Data to Drive Continuous Improvement." In *Beyond Transparency: Open Data and the Future of Civic Innovation*, edited by Brett Goldstein and Laura Dyson. Code for America Press. pp. 211-232.

Schein, Richard H. 1993. "Representing Urban America: 19th-Century Views of Landscape, Space, and Power." *Environment and Planning D: Society and Space* 11 (1): 7–21.

Schwanen, Tim, and Mei-Po Kwan. 2009. "'Doing' Critical Geographies with Numbers." *The Professional Geographer* 61 (4): 459–464.

Scott, David W. 1985. "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions." *The Annals of Statistics* 13 (3): 1024–1040.

Sheller, Mimi, and John Urry. 2006. "The New Mobilities Paradigm." *Environment and Planning A* 38 (2): 207–226.

Shelton, Taylor, Ate Poorthuis, Mark Graham, and Matthew Zook. 2014. "Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of 'Big Data.'" *Geoforum* 52: 167–179.

Shelton, Taylor, Matthew Zook, and Mark Graham. 2012. "The Technology of Religion: Mapping Religious Cyberscapes." *The Professional Geographer* 64 (4): 602–617.

Shelton, Taylor, Matthew Zook, and Alan Wiig. 2015. "The 'Actually Existing Smart City'". *Cambridge Journal of Regions, Economy and Society* 8 (1): 13-25.

Sheppard, Eric. 2001. "Quantitative Geography: Representations, Practices, and Possibilities." *Environment and Planning D: Society and Space* 19 (5): 535–554.

Silm, Siiri, and Rein Ahas. 2014. "Ethnic Differences in Activity Spaces: A Study of Out-of-Home Nonemployment Activities with Mobile Phone Data." *Annals of the Association of American Geographers* 104 (3): 542–559.

Söderström, Ola. 1996. "Paper Cities: Visual Thinking in Urban Planning." *Cultural Geographies* 3 (3): 249–281.

Stefanidis, Anthony, Amy Cotnoir, Arie Croitoru, Andrew Crooks, Matthew Rice, and Jacek Radzikowski. 2013. "Demarcating New Boundaries: Mapping Virtual

Polycentric Communities through Social Media Content." *Cartography and Geographic Information Science* 40 (2): 116–129.

Wall, Melissa, and Treepon Kirdnark. 2012. "Online Maps and Minorities: Geotagging Thailand's Muslims." *New Media & Society* 14 (4): 701–716.

Warf, Barney, and Daniel Sui. 2010. "From GIS to Neogeography: Ontological Implications and Theories of Truth." *Annals of GIS* 16 (4): 197–209.

Watkins, Derek. 2012. "Digital Facets of Place: Flickr's Mappings of the U.S.-Mexico Borderlands". Unpublished M.A. Thesis, University of Oregon Department of Geography.

Welch, Jack. 2013. "The Great Changeover". *Louisville Magazine* (March): 30-35.

Wilson, Matthew W. 2012. "Location-Based Services, Conspicuous Mobility, and the Location-Aware Future." *Geoforum* 43 (6): 1266–1275.

Wong, David W. S., and Shih-Lung Shaw. 2011. "Measuring Segregation: An Activity Space Approach." *Journal of Geographical Systems* 13 (2): 127–145.

Wyly, Elvin. 2009. "Strategic Positivism." *The Professional Geographer* 61 (3): 310–322.

Wyly, Elvin. 2011. "Positively Radical." *International Journal of Urban and Regional Research* 35 (5): 889–912.

Wyly, Elvin. 2014. "The New Quantitative Revolution." *Dialogues in Human Geography* 4 (1): 26–38.

Zook, Matthew, and Mark Graham. 2007. "The Creative Reconstruction of the Internet: Google and the Privatization of Cyberspace and DigiPlace." *Geoforum* 38 (6): 1322–1343.

Chapter 5
Conclusion

I. Summary of Dissertation Findings

The three papers in this dissertation have explored the possibilities for a relational socio-spatial analysis of geotagged social media, together providing an alternative framework for using these emerging sources of so-called big data for geographic research. By providing a counter to both those analysts who uncritically celebrate the potentials of big data in social research and those who critique the big data establishment for such naivety without actually engaging with the data itself, this dissertation has hoped to point towards a more fruitful marriage between geography and big data, reminiscent of earlier attempts within the discipline to bring together the tools and methods of GIScience with the theoretical and political commitments of its critics from human geography. In heeding Crampton et al's (2013) call to go 'beyond the geotag' in social media mapping, this dissertation has sought to combine the methods and epistemology of critical GIScience with the spatial ontology of relational socio-spatial theory, coming especially out of the work of Doreen Massey and Ash Amin. Together, this combination of existing, but heretofore largely unrelated, frameworks for understanding the world provides a significant counterpoint to those who see big data as representing a fundamental challenge to the concepts and methods of geography, or even of the social sciences writ-large. As such, this dissertation has pointed towards the significant possibility for engaging with these new sources of data outside of the confines of a naïve positivist epistemology, a strictly-quantitative methodology and a long-since superseded Cartesian spatial ontology. More specifically, this dissertation has sought to accomplish three key things as it relates to the development of a relational socio-spatial analysis of geotagged social media data.

First, this dissertation has sought to expand the conceptual grounding on which the original critique of the 'spatial ontology of the geotag' was based. Chapter 2 in particular explicates how more mainstream or popular social media mapping projects have mobilized overly simplistic understandings of space, which limit both the kinds of questions one can ask and the kind of insights one can draw from the analysis of this data, while also serving to dissuade more serious geographical and social-scientific scrutiny of this data. By situating the growth of these kind of limited analyses within longer histories of Cartesian spatial theory and GIScience, this dissertation points towards the possibilities contained within the utilization of alternative, relational understandings of space and spatiality developed within human geography in the past two to three decades.

Second, this dissertation has shown that despite the persistent Cartesianism inherent to many forms of geographic data and representation via geographic information systems, it *is* possible to operationalize a more multidimensional and relational understanding of space within the analysis of these new sources of data. In avoiding a whole host of social media data which doesn't contain explicit geographic information – which itself offers a range of opportunities to go 'beyond the geotag' in understanding more implicit geographic references – this dissertation has continued to mobilize the single pair of latitude and longitude coordinates attached to each individual data point, the kind of Cartesian spatial referent that allows for the geovisualization of this data in the first place. That is, while this work has sought to go *beyond* this Cartesianism, it hasn't done away with it altogether. But each individual point also contains the potential for being related to other points through a range of different methods of data collection, analysis and mapping. Chapter 3 in particular attempts to operationalize this understanding to show how one particular articulation of relational socio-spatial theory – Jessop et al's (2008) TPSN framework – can be applied to the mapping of social media data in order to uncover aspects of the data and underlying socio-spatial processes that are otherwise hidden by simply searching for concentrations of data points in a given locality. Chapter 4 similarly demonstrates how while these geographic coordinates are the basis for a socio-spatial analysis of this data, this data also allows for an 'explosion' of received, bounded and supposedly internally-coherent spatial categories.

Finally, it has been shown that a more relational analysis – combining qualitative, quantitative, spatial-analytic and cartographic methods – has the potential to reveal otherwise hidden stories that are concealed by overly simplistic approaches to this data, as well as more substantive insights into questions of urban socio-spatial inequality, as is shown in Chapter 4. While this data remains limited in its ability to speak definitively about social problems and processes, the application of a relational analysis to this data can allow for the creation of empirically-grounded counter-narratives to more conventional discourses about urban inequalities, which tend to rely on similarly simplistic understandings of geography and which in turn pathologize those who are the victims of unjust social structures.

A. Key Concerns for Geographic Research Using Geotagged Social Media Data

Through this exploratory approach, this dissertation identifies the following key concerns for future geographic research attempting to employ geotagged social media data from a relational perspective, from the formulation of a research question that can be addressed with this data through the interpretation of the data after its analysis. First, when thinking through the possibility of using social media data for geographic research, researchers should be aware of what social media represents and what kinds of research it enables. That is, this kind of data represents the empirical realities of everyday life for a large number of people, specifically their locational traces at particular moments in time and, to a lesser extent, qualitative information about their thoughts or feelings in that moment. The population represented in this data isn't a representative sample of society, however, and so this data is imprecise at best in helping to make broader claims or predictions about social processes. Instead, this data is most useful for analyzing individuals' movements through space over time and identifying connections between these places. That being said, nothing in this data allows for an identification of causal mechanisms or deeper underlying rationales for why the data is what it is; ultimately, this data is most useful as a way of documenting and analyzing the 'where' and the 'what' of socio-spatial processes, rather than the ‘how’ or ‘why’. Should one be more interested in these types of ‘how’ or ‘why’ questions, social media data is unlikely to be of much use for one’s purposes.

Second, should geotagged social media data be suitable for one’s research questions and objectives, the data collection process should emphasize a multiple step process of filtering data based on a variety of different attributes. That is, simply collecting a dataset of tweets mentioning a specific keyword, or a dataset of all tweets in a given area, is unlikely to provide substantive insight into any question, no matter how many individual points there may be. But by filtering this data based on multiple criteria, researchers can formulate more targeted subsets of data that group individual users based on different kind of topical of geographic affinities, thus providing a more relevant and substantial dataset that one can explore in more detail. Third, the process of analyzing this data should go beyond simply plotting these points on a map, and instead should emphasize putting this data into a broader context through processes of comparison. Whether this is through the relatively simple process of normalizing data by a random sample or comparing two ‘opposites’ against one another, or combining the analysis of social media data with other ancillary data sources, putting a given dataset into relation with other datasets is a key component of going ‘beyond the geotag’. Together, these two concerns allow for a more deliberate integration of relational socio-spatial theory, focusing on the interrelationships between different places and social processes, rather than using overly simplistic or received categories on which to base one’s analysis.

Finally, the interpretation of the data should draw explicitly on existing, topically relevant research and, in many cases, be supplemented by more conventional qualitative or quantitative methods. That is, because this data is strongest at describing empirical realities, it works best as a kind of supplement to existing work, either confirming theories through new forms of empirical documentation, or providing a new basis for which one could further interrogate the causal relationships that led to such an empirical reality. This point further emphasizes that the data should never be left to 'speak for itself', but is only given meaning through more-or-less subjective processes of interpretation.

II. Limitations of the Present Research and Potentials for Expansion

While this dissertation's exploratory emphasis has pointed to the potentials for the kind of relational socio-spatial analysis described above, this study has also demonstrated a number of shortcomings, even within such a relational framing. First, and perhaps most importantly, the conceptual and methodological focus of this dissertation ultimately meant that the dissertation as a whole was somewhat erratic from a topical or substantive point-of-view, moving from one issue to the next without a prolonged engagement with any single topic. Focusing on a single issue – such as the Louisville case in Chapter 4 – may have had the potential to yield much more substantive insights of that kind, rather than simply demonstrating how one might go about operationalizing a series of conceptual and methodological framings in relation to geotagged social media data. Though Chapter 4 points the furthest towards the potentials for this kind of research, it also demonstrates the necessity of embedding the analysis of social media data within a more prolonged engagement with a given issue, if one is to be able to make substantive claims based off of this kind of analysis. As such, future research utilizing geotagged social media data needs to strive for deeper, more substantive engagements with a single issue of social importance over a longer period of time. Similarly, the analysis of social media data ought to be combined with more conventional sources of data and methodologies – from archival research to interviews and ethnography or spatial analysis of more 'official' sources of social data.

The second major shortcoming of this dissertation, in some ways related to the first, is that the kind of in-depth qualitative analysis needed in relation to this data was never really demonstrated. Though each of the three papers in this dissertation include some element of looking at qualitative data in the body of tweets to accompany the statistical and cartographic analysis, this work was never particularly systematic. This is, of course, owed at least in part to the large number of individual tweets in the dataset, which can be quite unwieldy, and to the varying topical focus of each of the three papers, which prevented a more in-depth perspective necessary to make sense of the qualitative aspect. Even when looking at smaller subsets of tweets, such as the tweets about the 57th Street crane in New York City or other Sandy-related tweets in Los Angeles as seen in Chapter 3, it remains incredibly difficult to effectively interpret and summarize these many bursts of 140 characters (or, in some cases, less). So whether it be through long-term observations of tweeting about a single issue or the supplementing of tweet analysis with offline methodologies (e.g., interviewing the users represented in a given subset of data about their thoughts and motivations), there remains significant potential for geographers to add to the qualitative analysis of this data without relying on the still-problematic automated processes for sentiment analysis, or other more quantitatively-oriented approaches to dealing with qualitative data (cf. Burgess and Bruns 2012 for a similar call for in-depth textual analysis of social media data).

In addition to these two primary shortcomings, a number of other issues remain relatively underexplored within this dissertation, and offer possibilities for further extension of the methods discussed here. For example, building off of the analysis presented in both Chapter 2 and Chapter 4, it is expected that an extension of this kind of analysis – starting from a bounded area and demonstrating its spatial extensivity through the movements of people who live, work or socialize there – to other scales could be potentially fruitful (see below for further discussion of this method as applied to other topical concerns). For instance, in the Louisville case discussed in Chapter 4, the West End represents an amalgamation of several different neighborhoods, each with a relatively distinct history and social composition, which tends to be erased with the grouping of these areas into 'the West End'. And while the analysis of Chapter 4 attempts to develop a more nuanced understanding of the West End as anything but a internally homogeneous, bounded social area, the application of this method to each of the distinct neighborhoods within the West End could help to highlight the similarities and particularities of each of these places. Applying this method to larger scales – from the city, the urban region, state or administration region, or even the nation-state – could also allow for greater linkages with more network-oriented approaches to political or economic geography that emphasize connectivity across space.

This dissertation is also somewhat limited in its attempts to visualize relationality. While the kind of flow map presented in Figure 2.3 represents a somewhat more intuitive – albeit rudimentary – way of visualizing relationality as compared to the hexagonal maps, the explicitly relational aspect of this analysis has required a more thorough textual explanation to accompany these visuals. Improved methods of visualizing relationality would similarly be applicable beyond the study of social media, and would more clearly bring together critical GIScience and relational socio-spatial theory as applied to other subfields of the discipline.

Finally, although this dissertation has been focused on geotagged social media as one particular iteration of the emerging interest in utilizing big data, it has been limited in its singular focus on Twitter data. While Twitter has generally represented one of the more open social media platforms as far as data sharing is concerned, a range of other platforms make their data available, offering further potential for cross-referencing and combining different data sources for socio-spatial analysis. For instance, Instagram photos would offer new possibilities for visual analysis of photographs, while also offering a higher proportion of content with explicit geographic references, as compared with Twitter (Wortham 2012; see also Hochman and Manovich

2013). Indeed, relying too heavily on a single proprietary platform risks the possibility of being shut out entirely as these companies opt to sell their data to third-party brokers as a way of gaining financial solvency, rather than simply giving it away to whoever wants it and is technically-savvy enough to collect it (cf. Thatcher 2014).

III. Future Research Directions

While much of this dissertation has been directed at demonstrating the compatibility of two previously disparate conceptual and methodological approaches, Chapter 4 serves as a guide for how a relational socio-spatial analysis of geotagged social media data can be useful for understanding the dynamics of urban socio-spatial processes. Though Chapter 4 focuses on developing an empirically-grounded alternative understanding of segregation in Louisville, the approach of this dissertation could be similarly applied to other urban geographic topics, such as interrelated questions of territorial stigmatization, gentrification and neighborhood change.

Using the more simplistic approaches adopted elsewhere would seemingly indicate that such complex and socially-relevant issues are ill-suited for the application of geotagged social media data; collecting keywords like 'gentrification' or 'yuppie' would likely yield little in the way of substantive insight into these processes. It is both unlikely that these terms would be used in large enough numbers as to be significant, and even more unlikely that simple concentrations of tweets mentioning these or similar keywords would tell us anything about gentrification that wasn't already known. But by adopting the relational approach outlined in this dissertation, it is possible to understand processes of gentrification, like segregation, as fundamentally defined by the relations and connections between peoples and places and how these change over time, something that geotagged social media data is well suited for. As was mentioned above, geotagged Twitter data is useful for understanding how people inhabit and move through space. And by utilizing the previously demonstrated methods of filtering large datasets of tweets based on common characteristics, it is possible to go beyond simply tracking an individual's movements through the city to understanding how different groups of people use the city in different ways.

In order to empirically document processes of gentrification, researchers are typically limited to utilizing either intensive field surveys for identifying places of increased investment or census data in order to demonstrate changes in the class and/or racial composition of neighborhoods, data that is either difficult to produce and compare across cases or that is collected relatively infrequently and available only at predefined spatial scales that may not necessarily reflect the reality of the underlying processes of neighborhood change (Hammel and Wyly 1996; Wyly and Hammel 1998). But by understanding gentrification through the lens of a relational socio-spatial analysis, one could instead focus on how a given gentrifying area's connections with other spaces within the city have changed over time. Using geotagged tweets, one could map the spatial footprints of the users tweeting from a given gentrifying area, highlighting connections between the gentrifying area in question and other parts of the city where people that live in or frequent that area also spend time. By iterating this analysis at different time scales – for simplicity's sake, before and after a given watershed moment in an area's redevelopment – one could identify how these spatial footprints have changed through the process of gentrification, likely revealing a more spatially dispersed footprint as the neighborhood continues to gentrify, with more outsiders who had previously avoided the area due to processes of disinvestment and territorial stigmatization now coming into the area for consumptive activities, which tend to be well-recorded within social media datasets. So by understanding gentrification not only as a process whereby landowners are able to extract greater rents at the expense of the poor and racial minorities, but also as a process in which a neighborhood becomes more integrated with and connected to other spaces and places, one can complement the empirical documentation of this process through the use of geotagged social media data. Indeed, such an analysis could build off of the methods described by Hammel and Wyly (1996) for identifying gentrification, and work towards identifying the unique spatial footprint of gentrifying areas relative to other forms of redevelopment as seen through geotagged tweets. Even though this approach would allow for additional insight into the empirical realities of how gentrification plays out in peoples' everyday lives, this data is still unable to explain the 'why?' of gentrification; social media isn't capable of explaining how or why initial forms of disinvestment occurred, or who the primary actors engaged in promoting and profiting off of gentrification might be, or any number of other details that continue to best be answered through sustained qualitative engagement.

The more general approach offered by a combination of critical GIScience and relational socio-spatial theory is applicable to many of these same urban geographical questions, even beyond the use of geotagged social media data. Using more conventional data sources, such as local property data, can similarly benefit from a relational perspective by focusing not just, for instance, on the neighborhoods where vacant properties are located, but instead turning attention to those places where the *ownership* of these properties is concentrated. Applying such a relational approach to these questions produces alternative understandings of urban decline and territorial stigma, locating the production of these phenomena not in the neighborhoods themselves, but rather with 'outside' forces who are able to shape these neighborhoods in particular ways in order to maximize their own financial gains through absenteeism, speculation and concentrated ownership, among other things. Whether applied to new forms of digital data from social media platforms or longstanding forms of recording property transactions, this focus on analyzing and visualizing not just what Eric Sheppard (1995) called "the interrelations between attributes associated with specific locations", but also "the interrelationships between places" (11), offers significant potential for urban geographical research to reimagine spaces and places and rethink how both the problems facing cities, and solutions to such problems, rely on understandings of urban spaces as bounded, coherent and isolated from one another, rather than as fundamentally interconnected and co-constituted. Ultimately, the sheer size, variability and constant production of new sources of data means that much remains to be explored and understood about social and spatial processes through this data, so long as geographically-rich conceptual frameworks and research questions help to drive such exploration and analysis.

IV. References

Burgess, Jean, and Axel Bruns. 2012. "Twitter Archives and the Challenges of 'Big Social Data' for Media and Communication Research." *M/C Journal* 15 (5).

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb." *Cartography and Geographic Information Science* 40 (2): 130–139.

Hammel, Daniel J., and Elvin K. Wyly. 1996. "A Model for Identifying Gentrified Areas with Census Data." *Urban Geography* 17 (3): 248–268.

Hochman, Nadav, and Lev Manovich. 2013. "Zooming into an Instagram City: Reading the Local through Social Media." *First Monday* 18 (7).

Jessop, Bob, Neil Brenner, and Martin Jones. 2008. "Theorizing Sociospatial Relations." *Environment and Planning D: Society and Space* 26 (3): 389–401.

Sheppard, Eric. 1995. "GIS and Society: Towards a Research Agenda." *Cartography and Geographic Information Science* 22 (1): 5–16.

Stephens, Monica, and Ate Poorthuis. Forthcoming. "Follow Thy Neighbor: Connecting the Social and the Spatial Networks on Twitter." *Computers, Environment and Urban Systems*.

Thatcher, Jim. 2014. "Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data." *International Journal of Communication* 8: 1765–1783.

Wortham, Jenna. 2012. "Instagram Refreshes App by Including Photo Maps". *The New York Times Bits Blog*. 16 August. Available from: http://bits.blogs.nytimes.com/2012/08/16/instagram-refreshes-app-include-photo-maps/

Wyly, Elvin K., and Daniel J. Hammel. 1998. "Modeling the Context and Contingency of Gentrification." *Journal of Urban Affairs* 20 (3): 303–326.