GeoJournal (2017) 82:721–734
DOI 10.1007/s10708-016-9713-3

Spatialities of data: mapping social media 'beyond the geotag'

Taylor Shelton

Published online: 30 March 2016

Abstract As emerging sources of so-called 'big data' are increasingly utilized in order to understand social and spatial processes, so too have these new data sources become the subject of harsh criticism from more critically-oriented geographers and social scientists. This paper argues that one of the major issues preventing a more productive dialogue between critical human geographers and those already engaging in the mapping and analysis of these new data sources is around the ways that space and spatiality are conceptualized in social media mapping. As such, this paper draws on and extends earlier critiques of the 'spatial ontology of the geotag', in which the geographic analysis of geotagged social media data overprivileges the single latitude/longitude coordinate pair attached to each individual data point, often leading to the kind of simplistic mappings and interpretations prevalent today. The goals of this paper are two-fold: first, to demonstrate how the spatial ontology of the geotag is implicitly operationalized within mainstream social media mapping exercises, and how this understanding of space remains incongruent with existing conceptions of space drawn from human geography. Second, using the example of tweeting in the wake of the August 2014 killing of an unarmed African–American teenager by a police officer in Ferguson, Missouri, this paper demonstrates how a more geographically-situated analysis of this kind of data, inspired by relational or multidimensional conceptualizations of space, can yield alternative understandings of the social processes embedded in such data.

Keywords Big data · Critical GIS · Relational space · Social media · Socio-spatial theory · Twitter

T. Shelton (✉)
Georgia Institute of Technology, Atlanta, GA, USA
e-mail: taylorshelton@gatech.edu

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). This 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 its general popularity and near-ubiquity throughout much of society—social media 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-twentieth 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 conceived 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. Because the use of this data is increasingly the domain of non-academics and non-specialists with profit-maximizing motives, it is these individuals that are largely setting the agenda for how this data is and will be used moving forward. As such, this paper seeks to subject such 'amateur' or popular social media mapping exercises to critical scrutiny in order to understand how the dominance of this approach promotes particular understandings of social and spatial processes, and how this data might be analyzed otherwise.

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. Through the utilization of different techniques for normalizing and filtering this data, we can see a variety of alternative geographies that emphasize the very particular groundedness of social activity in places, as well as the connections between quite spatially distanciated places, as opposed to the more simplistic geographic readings often provided by this kind of data. Ultimately, 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.

Conceptualizing space and spatiality in social media mapping

As the broader 'spatial turn' in the social sciences and humanities has taken hold (Warf and Santa 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 conceptualizations of space mobilized across these disciplines remain tied to long-since eclipsed 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-h 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).

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), 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 March 2016, Rogers' original interactive map of Ferguson-related tweeting (see Fig. 1) has been viewed over 295,000 times, with links to the map appearing on roughly 300 other websites and being

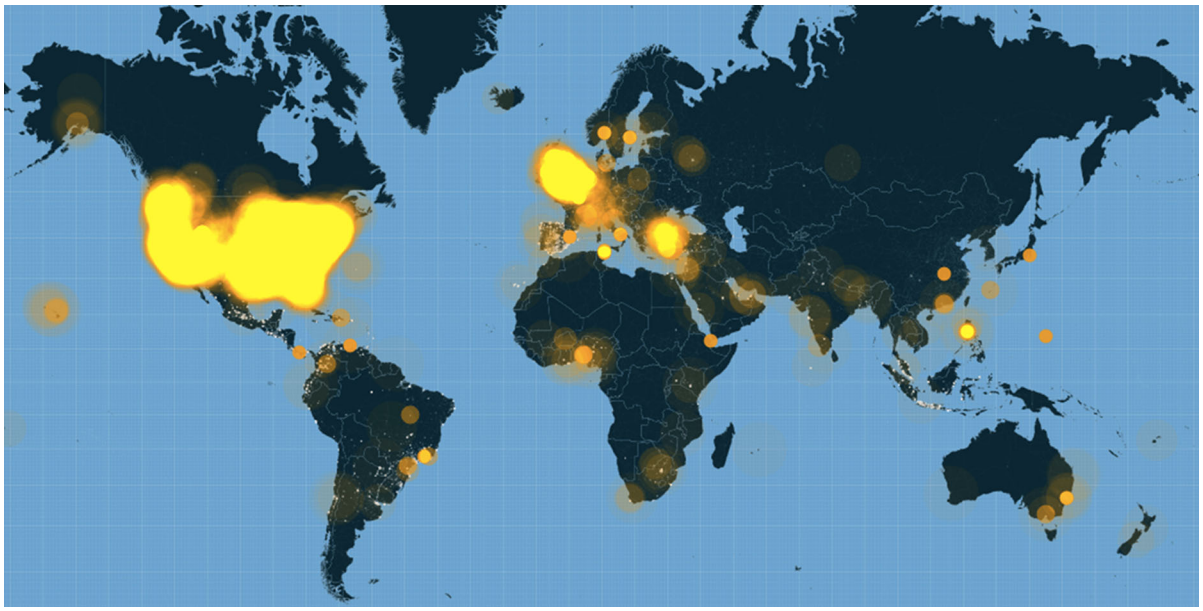

Fig. 1 "Tweets mentioning Ferguson" by Simon Rogers. *Source*: Author's Screenshot. Full map available from: http://srogers.cartodb.com/viz/4a5eb582-23ed-11e4-bd6b-0e230854a1cb/embed_map

shared over 1400 times via Twitter itself and engaged with over 1000 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 3700 tweets containing links to the map.[1]

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 point 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 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 data is decontextualized in both space and time, 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. 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

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

(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 spatially-indiscriminate. 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.

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 30 years of geographic thought have seen a dramatic shift towards much more complex and situated understandings of space and spatiality that stand 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 can tend 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 20 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.

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 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. For example, in their initial attempts to go 'beyond the geotag', Crampton et al. (2013) used the location of users retweeting a given tweet to identify the relational networks of Twitter users and how the diffusion of retweets reflects the particular social and spatial networks that users are embedded within. More recently, Shelton et al. (2015) have attempted to connect groups of users to those neighborhoods where they tweet the most, then using those identifications as ways of understanding how different groups of people move through and utilize the city in different ways. One key insight from Shelton et al's analysis is that due to the fundamentally mobile nature of contemporary social life, it's important that simple analysis of the presence or absence of user-generated geographic information in a given place be equated with the lack of engagement with these technologies. Indeed, even though predominantly poor and African-American neighborhoods may see less tweeting within their borders than comparable affluent and white neighborhoods, this is as much a function of the mobility of users from poorer neighborhoods, who are actually contributing a significant amount of the data being produced in wealthier neighborhoods elsewhere in the city. These insights about the utility of the other aspects of each data point, which allow for more complex ways of filtering data to reveal different kinds of social and spatial connections, inform the approach taken in this paper. 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.

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 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 (cf. Munroe 2012). 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. 2016, for more discussion of these methods, or Shelton et al. 2014, 2015 for examples of their application).

In order to demonstrate the utility of this approach, data was collected for a textual references to a variety of relevant terms in the 7 days following August 9, 2014 at 3 pm 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. This normalization was accomplished by calculating the odds-ratio at the lower bound of the 95 % confidence interval, using the following formula:

$$OR_{lower} = e^{\ln(OR_i) - 1.96 * \sqrt{\frac{1}{p_i} + \frac{1}{p} + \frac{1}{r_i} + \frac{1}{r}}}$$

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 95 percent confidence, that there are more points related to the phenomenon than one should expect, and vice versa for anything under 1. By calculating the odds-ratio, 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 Fig. 1, Fig. 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 4606 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.

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 Fig. 2, the 4965 Ferguson-related tweets from outside the United States were filtered according

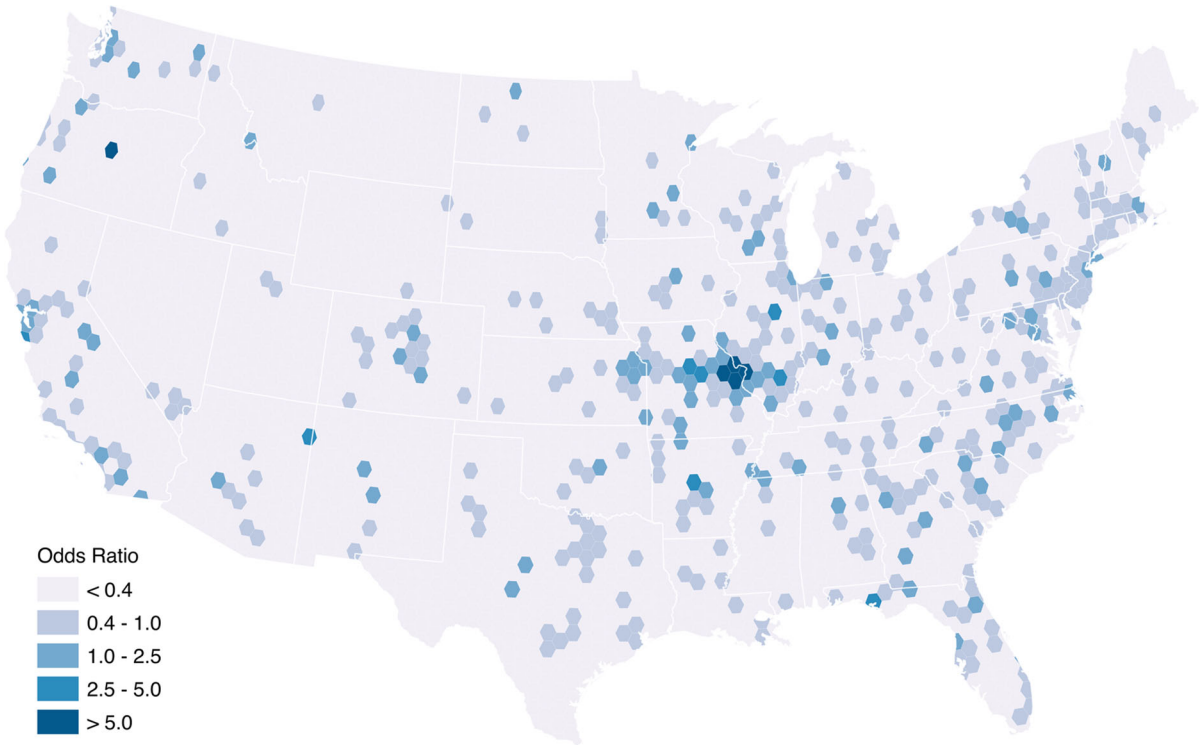

Fig. 2 Normalized map of Ferguson-related tweets in the United States

to the aforementioned other metadata fields to identify how many of the 2894 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, 1002, or more than a third, had also created tweets within the United States in the 3 year period between July 2012 and June 2015, totaling 678,078 geotagged tweets.

This data was then further filtered to identify those users who had previously tweeted from the area identified in Fig. 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 simplify 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 5119 tweets from within this area. Figure 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

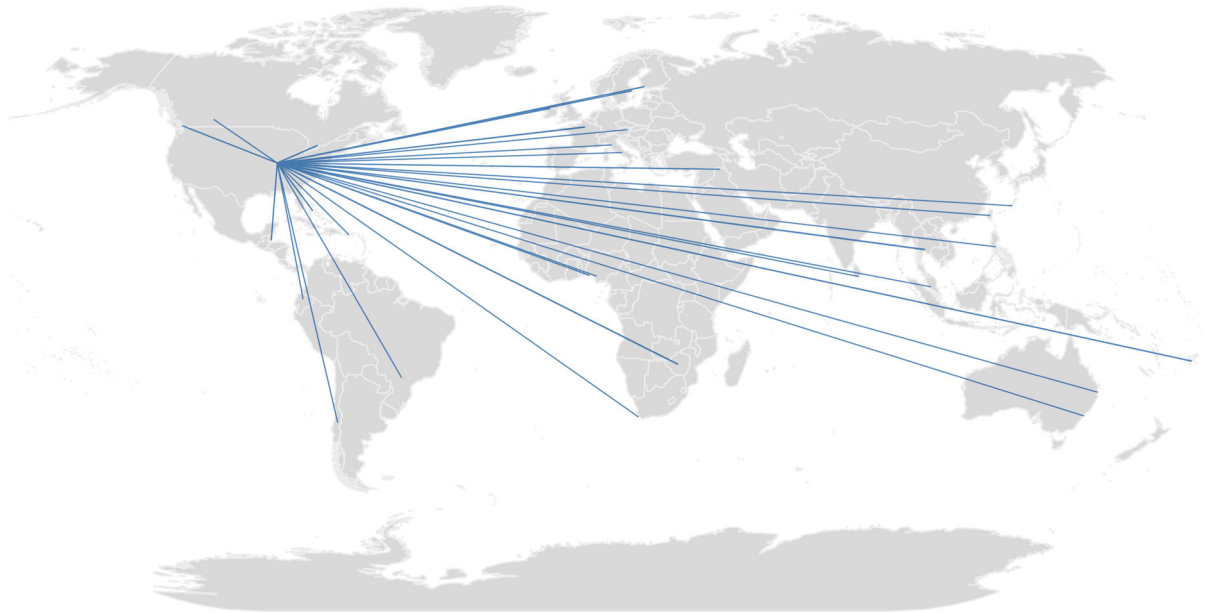

Fig. 3 Relational spaces of Ferguson-related tweeting

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 "2000 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.

And 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. Even though filtering down to the local level in this case means 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).

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 case of tweeting about the events in Ferguson, Missouri in August 2014. Rather than simply plotting points on a map in order to marvel at the interest in these events, we can apply a variety of statistical, cartographic and data collection methods in order provide a more nuanced understanding of the geography of reactions to Michael Brown's killing in Ferguson 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, 2015), 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. In particular, using the 'life histories' of Twitter users over time allow for a significantly greater amount of context to be added to the interpretation of these data points, especially as it relates to understanding the connectedness of seemingly disparate and distanciated places.

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 seek to more directly address substantive questions of inequality. 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.

References

Amin, A. (2002). Spatialities of globalisation. *Environment and Planning A, 34*(3), 385–399.

Amin, A. (2007). Re-thinking the urban social. *City, 11*(1), 100–114.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine 16*(7). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Barnes, T. J. (2013). Big data, little history. *Dialogues in Human Geography, 3*(3), 297–302.

Bonilla, Y., & Rosa, J. (2015). #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist, 42*(1), 4–17.

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

Brownlee, J. (2014). Watch how the twitter conversation of #Ferguson spread across America". *FastCo Design*. 18 August. http://www.fastcodesign.com/3034472/infographic-

of-the-day/watch-how-the-twitter-conversation-of-ferguson-spread-across-america.

Capps, K. (2014). Twitter made an amazing map of twitter going nuts over Ferguson. *CityLab*. 14 August. 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. http://blog.cartodb.com/cartodb-twitter-maps-in-today-yahoo-espn-time-cnbc/.

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

Cheng, Z., Caverlee, J., & Lee, K. (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. 759–68) CIKM'10. New York, NY: ACM.

Crampton, J. W. (2010). *Mapping: A critical introduction to cartography and GIS*. Hoboken: Wiley.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the Geotag: Situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science, 40*(2), 130–139.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (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, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #Earthquake: Twitter as a distributed sensor system. *Transactions in GIS, 17*(1), 124–147.

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

Curry, M. R. (1995). On space and spatial practice in contemporary geography. In C. Earle, K. Mathewson, & M. S. Kenzer (Eds.), *Concepts in human geography* (pp. 3–32). Lanham: Rowman and Littlefield.

Davis, C. A., Jr, Pappa, G. L., de Oliveira, D. R. R., & de L Arcanjo, F. (2011). Inferring the location of twitter messages based on user relationships. *Transactions in GIS, 15*(6), 735–751.

Field, K. (2013). The flawless map. *Cartonerd*. 17 December. http://cartonerd.blogspot.com/2013/12/the-flawless-map.html.

Field, K. (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. https://twitter.com/kennethfield/status/478775510386741248.

Fischer, E. (2010). *Locals vs. tourists*. https://www.flickr.com/photos/walkingsf/sets/72157624209158632/.

Fischer, E. (2014). Making the most detailed tweet map ever. *Mapbox Blog*. 3 December. https://www.mapbox.com/blog/twitter-map-every-tweet/.

Fung, B. (2014). Watch twitter explode along with Ferguson. *The Washington Post*. 14 August. http://www.washingtonpost.com/blogs/the-switch/wp/2014/08/14/watch-twitter-explode-along-with-ferguson.

Goodchild, M. F. (2006). GIScience ten years after ground truth. *Transactions in GIS, 10*(5), 687–692.

Goodspeed, R. (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*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6178.

Graham, S. (1998). The end of geography or the explosion of place? Conceptualizing space, place and information technology. *Progress in Human Geography, 22*(2), 165–185.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography, 3*(3), 255–261.

Graham, M., & Zook, M. (2011). Visualizing global cyberscapes: Mapping user-generated placemarks. *Journal of Urban Technology, 18*(1), 115–132.

Graham, M., & Zook, M. (2013). Augmented realities and uneven geographies: Exploring the geolinguistic contours of the web. *Environment and Planning A, 45*(1), 77–99.

Hargittai, E. (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.

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

Hecht, B., & Stephens, M. (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. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8114.

Hill, A. (2015). *In defense of burger cartography: Or, time to fall in love with maps all over again*. 28 March. http://andrewxhill.com/blog/2015/03/28/in-defense-of-burger-cartography/.

Kirsch, S. (1995). The incredible shrinking world? Technology and the production of space. *Environment and Planning D: Society and Space, 13*(5), 529–555.

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 2053951714528481.

Kitchin, R. (2014b). *The data revolution: Big data, open data, data infrastructures and their consequences*. Thousand Oaks, CA: Sage.

Kwan, M. (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, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global twitter heartbeat: The geography of twitter. *First Monday*. doi:10.5210/fm.v18i5.4366.

Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2014). Social media fingerprints of unemployment. *arXiv*. http://arxiv.org/abs/1411.3140.

Lotan, G. (2011). Data reveals that 'occupying' twitter trending topics is harder than it looks!. *Gilad Lotan Personal Blog*.

12 October. http://giladlotan.com/2011/10/data-reveals-that-occupying-twitter-trending-topics-is-harder-than-it-looks/.
Mahmud, J., Nichols, J., & Drews, C. (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, J., Nichols, J., & Drews, C. (2014). Home location identification of twitter users. *ACM Transactions on Intelligent Systems and Technology, 5*(3), 1–21.
Massey, D. (1991). A global sense of place. *Marxism Today, 35*(6), 24–29.
Massey, D. (1999). Philosophy and politics of spatiality: Some considerations. *Geographische Zeitschrift, 87*(1), 1–12.
Merriman, P., Jones, M., Olsson, G., Sheppard, E., Thrift, N., & Tuan, Y.-F. (2012). Space and spatiality in theory. *Dialogues in Human Geography, 2*(1), 3–22.
Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One, 8*(5), e64417.
Munroe, R. (2012). Heatmap. *xkcd*. https://xkcd.com/1138/.
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. https://twitter.com/NBCNews/status/537270894251749377.
Pew Research Center. (2015a). *Social media update 2014*. http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf.
Pew Research Center. (2015b). *Teens, social media and technology overview 2015*. http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf.
Poorthuis, A., & Zook, M. (2015). Small stories in big data: Gaining insights from large spatial point pattern datasets. *Cityscape, 17*(1), 151–160.
Poorthuis, A., Zook, M., Shelton, T., Graham, M., & Stephens, M. (2016). Using geotagged digital social data in geographic research. In N. Clifford, S. French, M. Cope, & T. Gillespie (Eds). *Key methods in geography* (3rd ed.). Sage: Thousand Oaks (forthcoming).
Rundstrom, R. A. (1995). GIS, indigenous peoples, and epistemological diversity. *Cartography and Geographic Information Systems, 22*(1), 45–57.
Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063–1064.
Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A, 38*(2), 207–226.
Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of hurricane sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum, 52*, 167–179.
Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning, 142*, 198–211.
Shelton, T., Zook, M., & Graham, M. (2012). The technology of religion: Mapping religious cyberscapes. *The Professional Geographer, 64*(4), 602–617.
Sheppard, E. (1995). GIS and society: Towards a research agenda. *Cartography and Geographic Information Systems, 22*(1), 5–16.
Sheppard, E. (2001). Quantitative geography: Representations, practices, and possibilities. *Environment and Planning D: Society and Space, 19*(5), 535–554.
Soja, E. W. (1980). The socio-spatial dialectic. *Annals of the Association of American Geographers, 70*(2), 207–225.
Stefanidis, A., Cotnoir, A., Croitoru, A., Crooks, A., Rice, M., & Radzikowski, J. (2013). Demarcating new boundaries: Mapping virtual polycentric communities through social media content. *Cartography and Geographic Information Science, 40*(2), 116–129.
Sui, D. (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, G. (2014). How facebook and twitter control what you see about Ferguson. *The Washington Post*. 19 August. http://www.washingtonpost.com/news/morning-mix/wp/2014/08/19/how-facebook-and-twitter-control-what-you-see-about-ferguson/.
Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of twitter networks. *Social Networks, 34*(1), 73–81.
Warf, B., & Arias, S. (2008). *The spatial turn: Interdisciplinary perspectives*. Abingdon: Taylor & Francis.
Widener, M. J., & Li, W. (2014). Using geolocated twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography, 54*, 189–197.
Wilson, M. W. (2015). Morgan freeman is dead and other big data stories. *Cultural Geographies, 22*(2), 345–349.
Wyly, E. (2009). Strategic positivism. *The Professional Geographer, 61*(3), 310–322.
Wyly, E. (2011). Positively radical. *International Journal of Urban and Regional Research, 35*(5), 889–912.
Wyly, E. (2014). The new quantitative revolution. *Dialogues in Human Geography, 4*(1), 26–38.