Aalborg Universitet

Soft City Sensing

A turn to computational humanities in data-driven urbanism

Madsen, Anders Koed; Grundtvig, Anders; Burgos-Thorsen, Sofie

Published in:
Cities

DOI (link to publication from Publisher):
10.1016/j.cities.2022.103671

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Madsen, A. K., Grundtvig, A., & Burgos-Thorsen, S. (2022). Soft City Sensing: A turn to computational humanities in data-driven urbanism. *Cities*, *126*. https://doi.org/10.1016/j.cities.2022.103671

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 04, 2025

Cities 126 (2022) 103671

Contents lists available at ScienceDirect

Cities

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

Soft City Sensing: A turn to computational humanities in data-driven urbanism

Anders Koed Madsen [a,*], Anders Grundtvig [b], Sofie Thorsen [a,c]

[a] Aalborg University Copenhagen, Denmark
[b] Copenhagen University, Denmark
[c] Gehl Architects, Denmark

ARTICLE INFO

Keywords:
Soft City Sensing
Urban studies
Big data
Social web
Digital city
Computational humanities

ABSTRACT

Data-driven urbanism is often entangled with the smart city and practiced in a way that prioritizes control over physical objects and downplays the human and political aspects of data. We label this approach 'hard city sensing' (HCS) and we argue that the rise of the 'digital city' offers the empirical foundation for more humanistic approaches. Driven by the ambition to untangle data-driven urbanism from HCS, this paper reviews two decades of scholarship that has used digital traces as an empirical ground for understanding urban phenomena. The review identifies four distinct ways of working with digital traces of which three pave the way for new ways of problematizing the city. Instead of abandoning the idea of data-driven urbanism, we propose the framework of 'soft city sensing' (SCS) as way to re-engage with it with inspiration from these pioneering works. However, this requires a willingness to revisit central epistemological commitments that currently serve as standards for how to "properly" do data projects. We therefore urge qualitative urban scholars to ponder the possibilities of furthering their urban interest by 'thinking with algorithms' while retaining their interpretative ambitions just as we identify a need for urban decion-makers to expand their criteria for what serves as valid data inputs to urban planning.

*The city as we imagine it, the **soft city** of illusions, myth, aspiration, nightmare, is as real, maybe more real, than the hard city one can locate on maps, in statistics, in monographs on urban sociology and demography and architecture.*

Raban (1974, our emphasis)

1. Introduction

What is a city? How do we get to know and improve it? The answer to these questions vary depending on disciplinary perspectives. To the traffic engineer the city could be a collection of movable objects to be controlled and efficiently managed. To the sociologist it could be a set of demographic patterns that make it possible to mitigate stratification. To the ethnographer it could be semiotic signs and interpretations that indicate how urban culture is weaved together. These are just three of many examples that illustrate an important point. 'The city' is an elusive phenomenon and we need a multitude of frames to make its different aspects tangible (Cukier et al., 2021). The fact that different disciplines have developed distinct conceptual and empirical approaches helps us view the city from many angles and retain its multiplicity. Fortunately, the field of urban studies has maintained an elective character that fits its object of analysis (Harding & Blokland, 2014). This ensures that a diversity of interests have a language and a toolkit through which they can couch their perspective on the urban. In the words of Kurgan & Brawley (2019) multiple 'ways of knowing cities' co-exist.

However, this eclectic character has proven difficult to translate into the executive rooms where decisions about urban futures are being made. Here, the city is often seen through standardized measurements and representations. Quantitative urban data—and the visualizations through which they are communicated—serve as the primary engines behind the professional vision of the people who have the power to define and solve urban issues (Goodwin, 2015). Historically, this is illustrated by Scott's (2020) exploration of how state bureaucracies in the modern era learned to see cities through technologies such as the census and bar charts. In our own era, we are similarly witnessing how

* Corresponding author.
E-mail address: akma@hum.aau.dk (A.K. Madsen).

https://doi.org/10.1016/j.cities.2022.103671
Received 18 March 2021; Received in revised form 11 February 2022; Accepted 17 March 2022
Available online 6 April 2022
0264-2751/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

sensors and interactive urban dashboards occupy a central role in the way proponents of smart city governance envision urban problems and their solutions (Kitchin, 2014b). Across these examples it seems fair to say that 'data-driven urbanism' has to a large extent become entangled with a distinct epistemology and philosophy of planning where data is seen as an instrument to predict and control large-scale urban infrastructures. This is what we loosely propose to call "hard city sensing" (HCS).

Our critical ambition in this paper is to use insights from computational humanities to *untangle* data-driven urbanism from this trajectory. The backdrop for this ambition is that the emergence of the 'digital city' (Halegoua, 2020) has resulted in a situation where urban activities leave qualitative and granular traces that affords data-intensive analyses with roots in humanistic epistemologies (Kitchin, 2014a). Analyses that are qualitative and explorative while simultaneously being data-intensive. We contribute to said untangling by critically reviewing two decades of scholarship that have pioneered the use of digital traces as an empirical foundation for producing urban insights and cartographies. We see these studies as indications of the way new data possibilities can productively be grasped within the field of urban studies. The review identifies four distinct ways of working with digital traces in an urban context. The approach we call 'sensor surrogates' does not take advantage of their qualitative and explorative affordances. The 'phenomenological grid' is an approach that takes advantage of the qualitative affordance, but not the explorative potential. The approach we call 'situated boundaries' takes advantage of the explorative affordance, but not the qualitative. Finally, we use 'Soft City Sensing' as a label for the papers that take advantage of both affordances simultaneously.

After having identified these approaches, we discuss the extent to which they pave the ground for new ways of problematizing the city and perhaps even stimulate a need for inviting new experts into data-driven urban planning. We argue that each of the three latter approaches carry such potential in their own distinct ways. Instead of critizising idea of data-driven urbanism from a diatance, we thus propose to re-engage with it with inspiration from these pioneering works. We use SCS as a headline for such reengagement which involves analyzing and framing the potentials of digital data in alternative ways than what the HCS toolkit offers. However, we also stipulate that any future institutionalization of SCS depends on a willingness to revisit central epistemological commitments that currently serve as standards for how to "properly" do data projects. Our hope is that this critical review and sketch for a future SCS agenda invite both qualitative urban scholars and urban decision-makers to rethink their stance on datafication. Whereas we urge the former to ponder the possibilities of furthering their urban interest by beginning to 'think with algorithms' while retaining their qualitative and interpretative ambitions, we ask the latter to expand their criteria for what serves as valid data inputs to urban planning.

2. Hard city sensing: data-driven urbanism under the smart city umbrella

We use the concept of data as a reference to inscriptions that allows selected aspects of the world to be analyzed and reorganized by computational techniques (Mayer-Schönberger & Cukier, 2013). We live in a world where this definition no longer just includes spreadsheets in Excel, but a plethora of images, texts, and relations on social media. This means that data-driven urbanism—which we define as the ambition to understand urban phenomena with reference to patterns in data—has the potential to engage professions and disciplines with urban interests and epistemologies that are rooted in the humanities. However, the dream of data-driven urbanism has historically been hard to disentangle from bureaucratic statecraft and it has recently manifested itself in the vision of the 'smart city' (Kitchin, 2017). While this concept is arguably vague and carries many connotations, Kitchin (2015) suggests that most of its incarnations share a dream with the bureaucracies described by Scott (2020). The dream of utilizing new data sources and algorithmic techniques to realize a more efficient and responsive form of urban planning. Often with an outset in predefined urban entities such as roads or administrative neighborhoods. According to critics of this development, the result has been that data-driven urbanism rests on three problematic assumptions about what data is and why it is useful in an urban context. We will touch briefly upon each in turn.

First, data and algorithms are reduced to tools for gaining *control* over complex physical urban systems (Caprotti, 2019). A central ambition of smart-city projects has been to model and predict urban processes, such as the flow of traffic or the potential breakdown of infrastructural systems (Batty et al., 2012). As a consequence, it has been central for data-driven urbanists to develop infrastructures of ubiquitous computing in the form, for example, of fiber cables, high-performance computers, and networks of urban sensors (Caprotti, 2019). In this technological infrastructure, data and algorithms are primarily used to track physical objects in space. Both academics and dedicated smart city initiatives from industry leaders - such as IBM and Siemens - have prioritized the development of IoT solutions to big urban problems like traffic, energy management, and public safety. Even a company like Google—which sits on a plethora of traces from the social web—takes a similar physicalist approach to its smart city initiatives in their Sidewalk Labs.

Second, this mode of data-driven urbanism results in analyses that lose track of the human aspect of the city. As put by Vanolo (2014), the resulting representations of urban space are characterized by "plenty of hi-tech symbols […] without any visible human presence" (Vanolo, 2014, 892). In line with this, Marvin and Luque-Ayala (2017) notes that most smart city projects are designed in a way that excludes stakeholders and viewpoints that could challenge the initial problem formulation by the project owner. As long as such dissonant voices are not inside the presumptions of the software used to represent the city, there is little chance that they will have an impact on the problematization of the city and the conclusions drawn about its development. The human aspects of the city are squeezed out of the operating systems in a way that risks reverting urban planning to ideals that are more reminiscent of the modernism criticized by Scott (2020) than the human-centered urbanism that has flourished in its wake.

Finally, as data-driven urbanism follows this path, it risks jeopardizing important democratic principles (Caprotti, 2019). The goals that predictive analytics are supposed to help achieve are often defined before problems and solutions from citizens are taken into account. Cardullo and Kitchin (2019) discuss this situation as a form of "civic paternalism" underpinning many smart city projects. Although such projects are often introduced as apolitical and non-ideological, they downplay political disagreement in a way that enables urban managers to steer their cities in their preferred direction (Kitchin, 2014a, 2014b, 2015). They carry interests beneath a mask of neutrality—a tendency that has also been discussed more boroadly under the heading of "technological solutionism" (Morozov, 2013).

We use the label 'HCS' as a loose reference to projects that frame the potentials of data with roots in these three assumptions, and we have just seen that urban scholars have during the last decades formulated relevant criticism of this powerful paradigm. However, whereas these critiques have been successful in exposing the opaque ideologies on which its rests, Kitchin (2015) have also pointed out that their authors have been reluctant to "undertake applied research aimed at **creating** smart city initiatives" (Kitchin, 2015: 134, our emphasis). For the most part, they have formulated their critique at a safe distance from data technologies without any interest in conducting experiments on how data-driven urbanism could be done differently. They rarely take advantage of the fact that our current data situation opens opportunities for alternative ways of measuring that could perhaps fix some of the problems identified. The result is a form of critical scholarship, which tends to be overly dismissive of the role of digital technology in urban planning and sometimes even indicates that being data-driven is in itself problematic (for important exceptions to this trend see e.g. Kurgan,

2013 and Williams, 2020). As mentioned in the introduction, we introduce SCS as a more interventionist line of critical scholarship (Zuiderent-Jerak, 2015) that draws on thoughts from computational humanities to actively intervene in the way data and algorithms are imagined and used in an urban context.

3. Disentangling data-driven urbanism: a turn to computational humanities

We use the concept of "computational humanities" as an umbrella term for the work of scholars who have, during the last several decades, used digital data sources and algorithmic techniques to create new methodologies for qualitative and exploratory analyses. Under headings such as "digital methods" (Rogers, 2013) and "digital humanities" (Berry, 2011), these scholars have especially seen empirical potential in the digital traces accompanying the rise of the social web. Whereas the smart city has prioritized a sensor-based data infrastructure that is optimized to model and predict the flow of objects in urban space, computational humanities have demonstrated how we can gain insights into human dynamics and sensemaking through digital traces such as pictures, reviews, and "likes" on web-based platforms. Rather than seeing data and quantification as being in opposition to the study of humans and culture, scholars in this field have used the "computational turn" (Berry, 2011) as an opportunity to invent a type of data science grounded in the epistemology of their own humanistic mother disciplines. In this way, they have attempted to take seriously the fact that many of digital traces were intentionally produced to convey meaning by their authors.

Envisioning data-driven urbanism from the perspective of computational humanities therefore opens for a different form of critical engagement with the HCS framework than the distanced critique discussed above. Rather than unpacking the hidden ideologies inherent in dominant versions of data-driven urbanism, the interventionist strategy we propose is to collect new types of urban data traces and produce urban representations that support alternative formulations of urban problems and thereby a different mode of planning. We motivate this proposition by highlighting two reasons why scholars working in the computational humanities have found the new data environment to have potential for their disciplines.

3.1. Reason 1: quali-quantitative traces and semantic techniques

Although it has been customary to make an equation between data and the quantitative sciences, many data sources are now simultaneously qualitative and quantitative (Venturini & Latour, 2010). A commentary thread on Facebook contains qualitative text and quantitative measures, such as the amount of likes on a given comment. An image on Instagram is a qualitative whole that consists of quantified pixels open for algorithmic investigation. Furthermore, the open source community has developed accessible algorithms that can be used to detect patterns in large volumes of quali-qualitative data. Techniques such as natural language processing make it possible to engage in "distant reading" (Moretti, 2013) of large corpora of texts, and developments in computer vision open the possibility for automatically identifying objects and people across large sets of images. Accordingly, algorithms can be used to conduct data-driven studies of how people perceive the world and the symbolic forms in which they communicate.

Goodchild (2007) indicated quite early that these types of traces could serve as an empirical foundation for cartographies of urban space that do not reduce people to physical points in space. People rate the urban environment on Yelp, depict it on Instagram, and indicate their willingness to engage in it on Facebook events. Because the symbolic systems of such digital platforms matter to the way we understand and navigate cities (Halegoua, 2020), they are valuable sources for understanding how urban life is made sense of by those participating in it. However, it has not been a priority for data-driven urbanism to realize this epistemic potential. In the years where computational humanists turned their gaze to traces of the social web, Townsend's (2013) influential book on smart cities explicitly dismissed these same traces as banal chatter and argued that at some point "the Internet of people gave way to the Internet of Things" (Townsend, 2013, 3). Rather than discussing the potential of qualitative traces, Townsend's (2013) book is filled with examples of how the possibility of tracking physical urban infrastructure—from snowplows to sewers—can provide useful feedback to urban planners and citizens. While maintaining a reflexive and critical stance toward this development, Townsend (2013) equates data-driven urbanism with the type of HCS discussed above.

This equation has grown so strong that people working on urban data have to a large extent missed the opportunity to experiment with the ways in which new forms of quali-quantitative data can make human life legible for urban planners. We think of the suggestion to turn data-driven urbanism toward computational humanities as an attempt to pick up a torch that was already lid in the 1960s and 1970s, where Whyte (1980) invented empirical techniques to record public life in New York City Streets and Gehl (2011) began measuring life between the buildings in Europe. Also Lynch's (1960) early experiments with mental maps carried a similar ambition of making cartographies of the human city. As digital traces stem from people's experienced reality, they can be used to map similar subject-centered itineraries that mark key features of the city from below. The goal is to tune into urban life rather than to model, predict, and control it. While this could be done exclusively through established qualitative methods, data-driven techniques from the computational humanities can provide a useful supplement in this endeavor. While such techniques cannot replace existing qualitative methods, they can guide qualitative analyses in a way that is similar to the way early ethnographers used map overviews to select and foreshadow their field sites before they ventured out in the world (Munk & Jensen, 2015). In short, big urban data can be immensely qualitative.

3.2. Reason 2: granularity and inductive algorithms

A second reason why scholars in computational humanities have found potential in the digital data environment is the possibility of combining granular data traces and techniques for pattern recognition into explorative tools (Nelson, 2020). While quantitative analyses have usually consisted of applying regression models to test pre-defined hypotheses, the possibility of using machine learning to underpin a more explorative data science has generated an interest among scholars used to work based on "grounded theory" or similar approaches. Rather than being constrained to divide humans into established classifications such as gender, age and race, digital data have opened the possibility for conducting so-called post-demographic studies where classifications emerge from patterns in data, rather than guiding the way they are analyzed (Rogers, 2013).

In an urban context, this approach could be translated into using a combination of granular geo-tagged data and explorative techniques to identify interesting urban areas and boundaries. For instance, if there is a tendency for events in a certain area to attract the same Facebook users, one could draw a data-driven polygon around that area and study it in detail. This would be an alternative approach to the widespread tendency to start spatial statistics from pre-defined urban grids such as streets, ZIP codes, elective districts, or other spatial divisions that often serve as the unchangeable reference against which interesting variations in data is observed. The idea that urban planning should take departure from such recognizable units has been a longstanding staple in the profession (Dewey, 1950; Galster, 1986), and it was explicated as a planning principle in the charter for New Urbanism two decades ago (Congress for the New Urbanism, 2000). However, this way of a priori choosing the relevant grid severely diminishes the flexibility of potential problem formulations.

Although urban statistics have usually not been conducted in such an explorative manner, we posit that the explorative potentials embraced

by computational humanities could also provide new empirical foundations for established theoretical trajectories in urban studies. For instance, the 'disticts' Lynch (1960) ended up identifying with his narrative methods did cross administrative boundaries. Also, in his classic text, "The City is Not a Tree," Alexander (1968) suggested that one of the problems of urban theory was its lack of ability to understand urban space as a set of dynamic and potentially overlapping units. While acknowledging the motivation to work with pre-established and mutually exclusive grids, he illustrated how many relevant aspects of urban life could only be grasped when the cartographer freed herself from predefined grids and embraced the ambiguity and situatedness of spatial boundaries. More recently, Kitchin and Dodge (2007) introduced the concept of the 'ontogenetic city' to make the similar point that space is not a container with pre-given attributes but a phenomenon that gains its form, function, and meaning in practice. We interpret this as a call for working with *softer* and more malleable urban boundaries and we argue that the combination of granular digital traces and inductive techniques enables us to pursue this agenda further.

4. Soft City Sensing: charting the contours of an alternative urban data paradigm

Our proposition is that data-intensive analyses with boots in the epistemic soil of the computational humanities can help make visible important aspects of urban life that fall outside the scope of HCS. We tentatively label such analyses 'SCS' to indicate that they (a) move beyond a physicalist ontology by attempting to inscribe the softer semantic aspects of urban life (e.g., people's sensemaking, experiential realities, and normative positions) and (b) refrain from having problem definitions dictated by pre-established urban grids and classifications and actively use the granularity of data sources to define different areas as relevant for different urban issues. However, we are by no means the first to suggest that the emerging quali-quantitative digital data environment carries potentials for such urban analyses. For instance, Williams (2020) recent call for 'data action' in urban studies makes a similar point as ours. We therefore introduce the framework of SCS with an outset in a review of academic works that have, during the last two decades, turned to the social web as an empirical source for studying and representing urban life. These studies show us the contours of a type of data-driven urbanism that have the possibility to realize the qualitative and explorative affordances of digital traces. The question is whether they do so and how?

4.1. Review method

Our review contains a close reading of 45 papers that we selected for review through complementary strategies. Our main strategy was to search SCOPUS for academic papers that employed digital traces from social media as an empirical ground to analyze urban life.[1] We used this SCOPUS search in combination with a snowballing strategy where we followed the references and links from papers that we thought illustrated SCS potentials. We read the titles and the abstracts, chose the most relevant articles and added them to our corpus of texts to review. Finally, we used the "cited by"-function at Google Scholar, that shows which articles that are referencing back to the source article. By doing so, we snowballed the other direction in time and now only received articles that were newer than the source article. About half of the papers came from the Scopus search and the other half from the snowball method. While not exhaustive, the 45 papers in our review provide a broad overview of how data from the social web have been reappropriated in urban studies since it gained widespread popularity around 2009. Fig. 1 shows the data sources used in the reviewed papers.

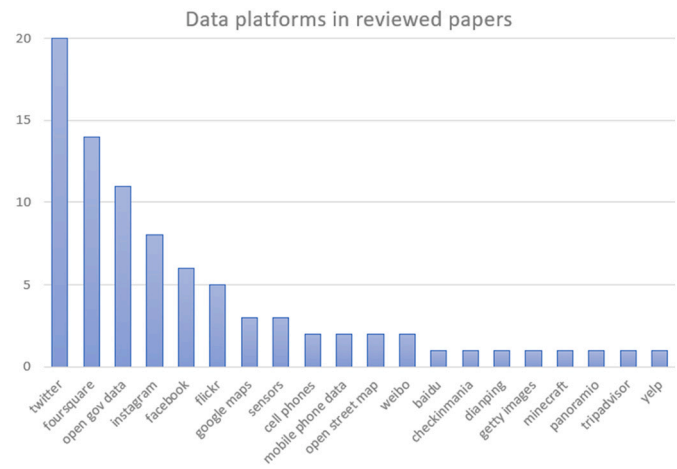

Fig. 1. The data sources used in the 45 reviewed papers and the number of papers using them. Note that a paper can use more than one platform.

Two meta-observations of the corpus are especially noteworthy. First, there is a bias toward Western and Anglo-Saxon platforms. Even though data from social media sites such as 微博 (Wēibó) and 微信 (WeChat) are used in a few of the articles, these platforms are not well represented in our sample. Second, a majority of data in the reviewed papers came from Twitter, Foursquare, and Facebook (we return to the open gov data later). This finding points to a text-centric bias in the reviewed papers. Also, the studies in our review sample that work with data from visual platforms such as Instagram or Flickr often do not make analytical use of the images. This under-prioritization of visual platforms and data in our reviewed papers mirrors broader trends in the literature pointed out by other scholars (see e.g., Highfield and Leaver (2016) and Thorsen and Munk (forthcoming)).

Our review analysis consisted in plotting the 45 papers on two dimensions that each mirrors the two reasons why scholars in the computational humanities have taken an interest in digital traces. We think of these as two affordances of digital data that are important to take advantage of to practice SCS. The vertical dimension in Fig. 2 below concerns the extent to which projects use traces from the social web to inquire into the *qualitative and semantic* components of urban life. Projects scoring 1 on this dimension use such traces as a surrogate for sensor data. If the answer to the question 'could this study have been done with an RFID sensor or a static camera?', was 'yes', then we placed the project low on the vertical dimension. An example could be the use of geolocations of tweets to study how people move in the city without regard for the semantic components of the tweet. Project scoring higher on this dimension are, to the contrary, interested in the lived life signaled by the digital traces. In such projects, the traces left by people through engagement with a digital platform could never be substituted by an "objective" sensor or camera. Insights into the intention behind the traces and the interpretation of the motivations of the urban agent are core to these studies. Such insights are achieved either through "thin" traces such as clicks and social buttons (e.g., likes and tags) or through "thick" traces, such as pictures and textual material. The latter projects score 3, while the former score 2. The horizontal dimension on Fig. 2 concerns the extent to which the papers use the granularity of digital traces to *redraw situated urban boundaries*. Projects scoring 1 on this dimension organize and interpret digital traces with reference to a pre-existing urban grid, which could be made up of administrative units like zip codes or more physical divisions such as streets or rivers. Projects scoring 3 on this dimension use spatial patterns in digital traces to redraw situated urban boundaries with relevance to the phenomenon studied. They take such boundaries to be fluid and they can be drawn with an outset in traces that are social, aesthetic, or based on sheer movement. In the middle of these positions are the projects scoring 2,

[1] The Scopus search query was: "*TITLE-ABS-KEY (("social media" OR "user generated content") AND ("city" OR "cities" OR "urban" OR "neighborhood"))*".

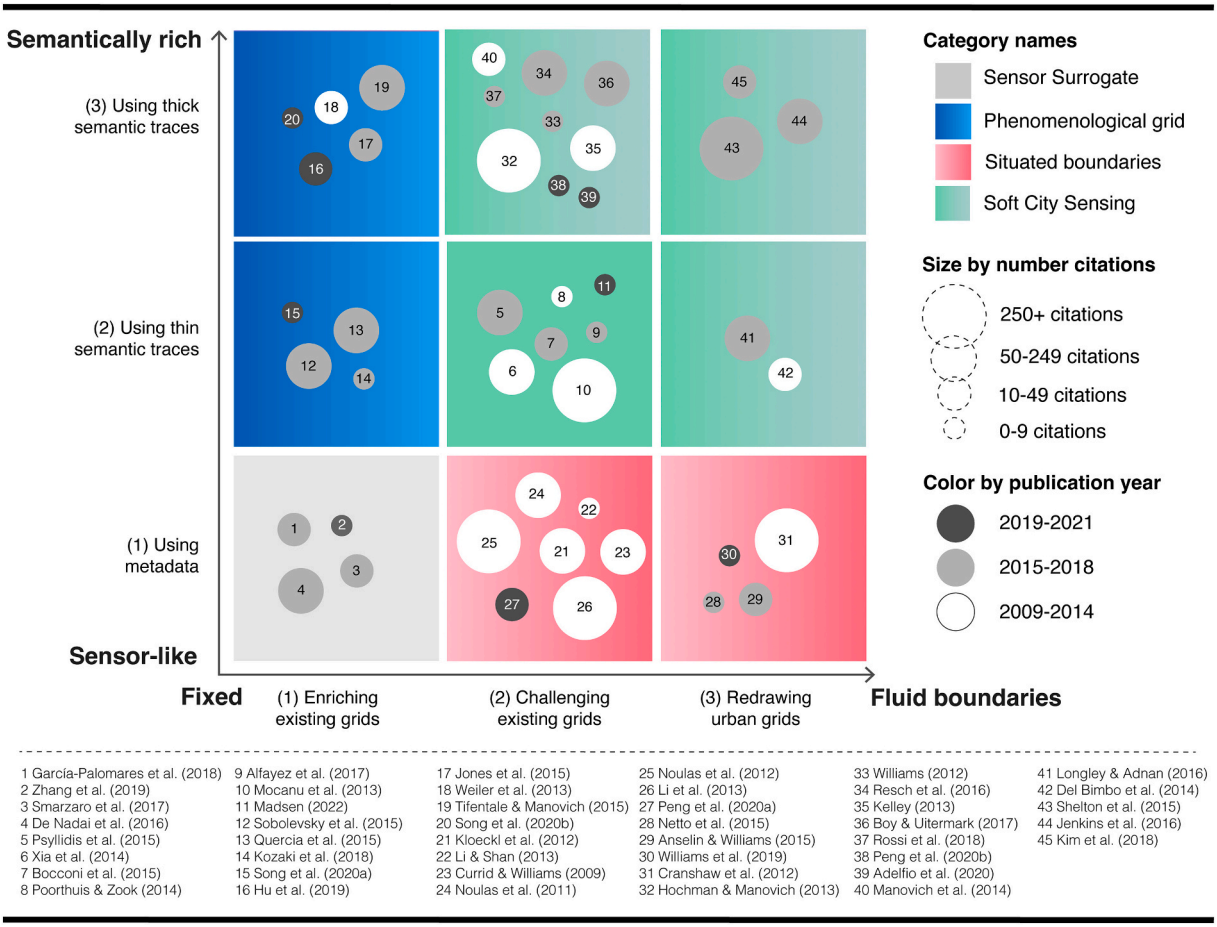

Fig. 2. The two-dimensional typology guiding our lit review. Forty-five papers were plotted based on the extent to which they took advantage of the phenomenological and granular affordances of data from the social web. The references to the papers are listed at the bottom of the image.

which criticize the relevance of existing grids without actively producing new ones themselves.

In the process of placing the papers on the two dimensions, each of us coded a selection of papers individually without consulting each other. To perform the coding, we read the papers with special attention to the methodology section, the data descriptions as well as the images and visualizations reported in the papers. Then, we placed the papers on the two dimensions and discussed differences between our coding results. In this process, we revisited and reframed the typology and developed the mode of scoring outlined above. Fig. 2 shows the final distribution of the reviewed papers and the four distinct ways of working with digital traces, to which our review gave rise. One comprises a small group of papers we call "soft sensor surrogates" (gray). They work with data from the social web without taking advantage of any the two affordances. We also identified two ways of working with digital traces that only took advantage of one of the two affordances that guided our review: "the phenomenological grid" (blue) utilized the qualitative potential in digital traces whereas "situated boundaries" (red) utilized their granularity. Finally, we identified a group of papers that took advantage of both affordances, and we labeled these "SCS" (green). Below, we describe these four distinct ways of using data from the social web and subsequently discuss the extent to which they open new ways of representing and problematizing the city.

4.2. Sensor surrogates: keeping new data on the beaten path

Sensor surrogates denote a group of studies that translate digital traces into indices of the physical city that could also have been obtained through sensors or surveillance cameras. The *intentions* and *experiences* behind the data points are not in focus. Furthermore, these studies make the choice to aggregate data on predefined spatial units, such as blocks (De Nadai et al., 2016), land zones (García-Palomares et al., 2018), city regions (Smarzaro et al., 2017; Zhang et al., 2019), or counties (Li et al., 2013). The methodological choice is to refrain from using any of the two affordances of data from the social web to move out of the two dimensions in our typology. Despite working with the social web as their data source, these studies carry on within a frame that is remiscent of HCS rather than exemplifying a 'turn to computational humanities'.

García-Palomares et al.'s (2018) study of city dynamics through Twitter is a paradigmatic example. To study the link between land use and urban activity in Madrid, the authors collected three million geotagged tweets to build a map that shows how some predefined zones, such as parks and retail areas, maintain constant activity throughout the day, while others fluctuate more. The authors explicitly states that their empirical strategy is to use tweet densities as a *surrogate* of population densities, which they aggregate on a grid borrowed from the Madrid transport authority (see fig 3 below). Despite acknowledging their semantic and qualitative potential, tweets are deliberately translated into moving points on a map. They are reduced to a proxy for people's position in the city. The humans behind the tweets are conceptualized as physical bricks moving in a recognizable spatial grid.

We found three other sensor surrogate papers in our review. De Nadai et al. (2016) used social media activity to map activity in urban venues, and Smarzaro et al. (2017) reduced online reviews to chart the availability of different types of services across the city. More surprisingly, Zhang et al. (2019) also falls in this category. Even though the

authors explicitly state the potential of using images from the social web to study people's subjective preference for the cityscape, this potential is not realized in the study. Instead of interpreting the intentions behind images, the paper uses machine learning to detect the objective physical characteristics that make a place visually unique. Characteristics that could just as well have been studied through images from a surveillance camera.

4.3. The phenomenological grid: from syntax to semantics

The nine papers falling under the heading of 'the phenomenological grid' all take advantage of the quali-quantitative affordances of digital traces while plotting these data within an established urban grid. These papers move up the vertical axis by taking advantage of either *thin* or *thick* semantic traces but stay low on the horizontal axis. In the thin traces category, Quercia et al. (2015) measured the safety and walkability of streets in Central London by counting Foursquare and Flickr activity. Without using the full qualitative potential of these data sources, the authors used metadata such as the gender of the user and the tags on the picture in their mapping. (Quercia et al., 2015). However, the richness of the user-generated material and the agency or intention behind each trace left by a user were not used to generate insights about their lived experiences. Adding more phenomenological nuance, Hu et al. (2019) used neighborhood reviews to perform a semantic analysis of the way different neighborhoods are perceived and problematized by their visitors (including issues of safety, cultural diversity, and life convenience). As seen in Fig. 4, the result is maps of New York City neighborhoods, which we think of as a contemporary version of mental and cognitive mapping.

Notwithstanding their differences, the phenomenological grid and sensor surrogates share the important trait of aggregating data within standardized spatial units. One consequence of this choice is that their visual outputs are—on the surface—very similar. The two maps in Figs. 3 and 4 are color-coded choropleth maps that compare predefined city zones. While the motivation for doing so, in the phenomenological grid papers, is to allow comparison between social media and other urban data collected within the same spatial units, it bypasses the chance of investigating how social traces might offer alternative and issue-specific ways of drawing city boundaries.

4.4. Situated boundaries: redrawing boundaries to fit urban issues

While papers in the phenomenological grid stay within established spatial divides to experiment with semantic mapping, the 11 papers in the situated boundaries category do not take advantage of the qualitative affordances of digital traces, but instead use their granularity to challenge standard urban boundaries. For instance, Cranshaw et al. (2012) used an algorithmic analysis of patterns in check-ins from Foursquare and Twitter to map the lived life in Pittsburgh. As shown in Fig. 5, the authors identify "livehoods" (red and blue colors) that cross existing administrative neighborhood borders (the black lines). For instance, the administrative neighborhood of Shadyside is argued to comprise two distinct clusters of urban life, with the western part dominated by older, richer people, and the eastern part dominated by younger, indie-looking people, whos urban life have a lot in common with the the adjacent neighborhoods of East Liberty.

While this map redraws urban boundaries and thereby exposes how existing administrative grids fail to fit the issue studied, the study does not dive into the qualitative content of the underlying data when describing these livehoods. The blue and red clusters are based on check-in patterns. They represent movements and co-locations between humans and could have been created using WIFI triangulation or other forms of granular tracking. While the authors do use these patterns as an outset for doing interviews, the qualitative methods have the character of an add-on to the geographical patterns. Other papers follow a similar strategy. For instance, Netto et al. (2015) used patterns in geo-coded tweets to produce proxies for actual trajectories that people follow when moving through the city. They used these data-driven trajectories to argue that administrative residential zones are not useful units when attempting to understand social segregation in Rio. Rather than conceptualizing segregation as something pertaining to residential neighborhoods, the authors re-ontologized the city as a network of encounters that pointed to alternative planning problems that cannot be properly understood through the established grid (Netto et al., 2015).

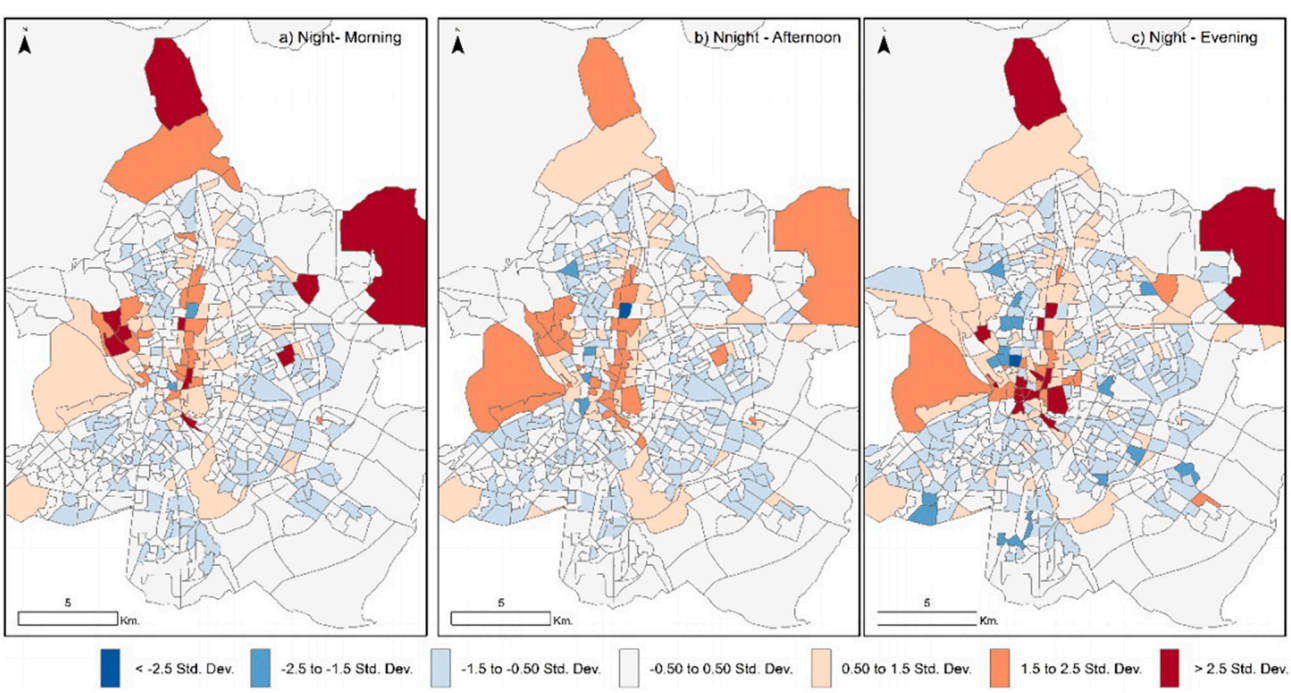

Fig. 3. Predefined areas in Madrid are colored based on the intensity of tweet activity throughout the day. The map is used to inform urban planners about the dynamic "pulse" of the city (by García-Palomares et al).

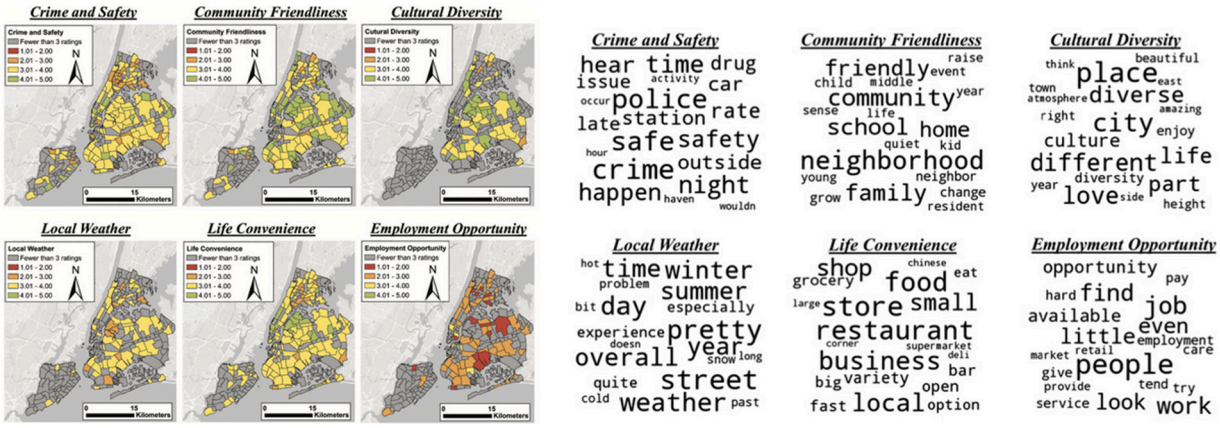

Fig. 4. The maps on the top use semantic patterns in neighborhood reviews to illustrate how people associate specific areas of New York City with specific urban issues. The word clouds below illustrate the words they use to describe these issues (by Hu et al).

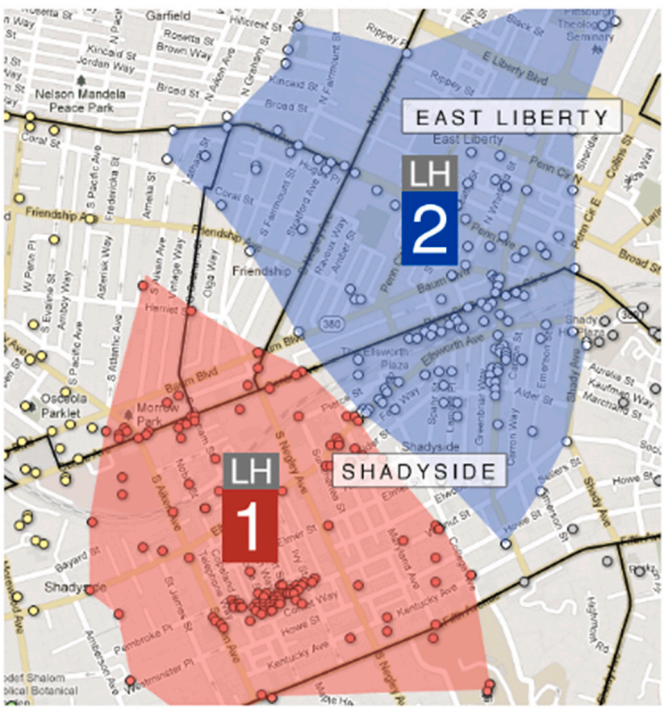

Fig. 5. This map plots check-ins from Foursquare to illustrate how "livehoods" (red and blue)—areas in which specific types of life unfold—cross administrative divisions (black lines) in Pittsburg. (by Cranshaw et al).

4.5. SCS: exemplars of an emerging research agenda

We use the label SCS to denote the papers that simultaneously move out on both dimensions and thereby take advantage of both the affordances that have been highlighted as interesting by scholars in the computational humanities. A paradigmatic example is a study by Shelton et al. (2015), who used geo-tagged tweets to visualize the *mental* boundaries through which people in Louisville make sense of their city (Fig. 6). More specifically, they did so by analyzing the semantic and geographical patterns around the use of the word "ghetto" in said tweets. The map uses data from the social web to portray a semantic urban space, and it deliberately frees itself from pre-existing spatial grids when delineating "the ghetto." Contrary to ambitions in HCS projects, this study does not solve the problem of efficiency or prediction, but rather one of prejudice. It is built with the deconstructivist purpose of disproving the living urban myth that 9th Street is a relevant boundary separating the east and west of Louisville. This finding can inspire local urban planners to rethink the identity of the city and pave the way for new urban narratives about its composition.

Another example of SCS is the way Jenkins et al.'s (2016) used a semantic analysis of textual data from Twitter and Wikipedia to identify areas in Midtown, New York City, that carry distinct atmospheres of entertainment. In line with the Louisville paper, the resulting map of New York serves to deconstruct established modes of thought. It challenges the idea that place atmospheres are geographically linked to physical venues (such as theaters), and it uses this insight to challenge a mental boundary that has guided destination management in New York for decades.

5. The kind of problem a city is… and could be!

In the previous section, we mapped the reviewed articles according to the way they empirically and analytically make use of social web traces to inquire into urban problems. In this section, we dive deeper into each of the identified typologies and unfold the arguments and propositions they make about how these data can be used to study cities. More importantly, we also outline the epistemological consequences of these commitments for how the city can be framed as a research problem and who can participate as urban experts within the given frames.

First, the sensor surrogate papers illustrate that the two affordances of social media data explored here—granularity and semantic richness—are not necessarily leveraged in a way that translates into reformulations of urban problems. By reducing the city and its inhabitants to physical entities, the sensor surrogate papers center on the same when-and-where questions about how people move about in the city that characterizes HCS. These studies do not open new urban problem formulations or involve new types of experts in their solutions, nor do they suggest ways to escape or reformulate what Kryssanov et al. (2001) once called the "autopoietic city". A city where humans are just one of many relevant physical objects to model and control from a bird's-eye view (de Waal, 2017). To be fair, the possibility of alternative problem formulations is not at all the reason why these studies turned to the social web for empirical data in the first place. The reason is rather that the authors found established methods too "slow and expensive" (Smarzaro et al., 2017: 1463) or "hard to scale up" (Zhang et al., 2019: 2). In these studies, the fascination with social media traces is framed as the possibility of harvesting a lot of data quickly and cheaply. This also explains why, for example, Li et al. (2013), De Nadai et al. (2016), and Smarzaro et al. (2017) make the methodological choice to source their data across multiple platforms. While this choice limits the possibility of diving into

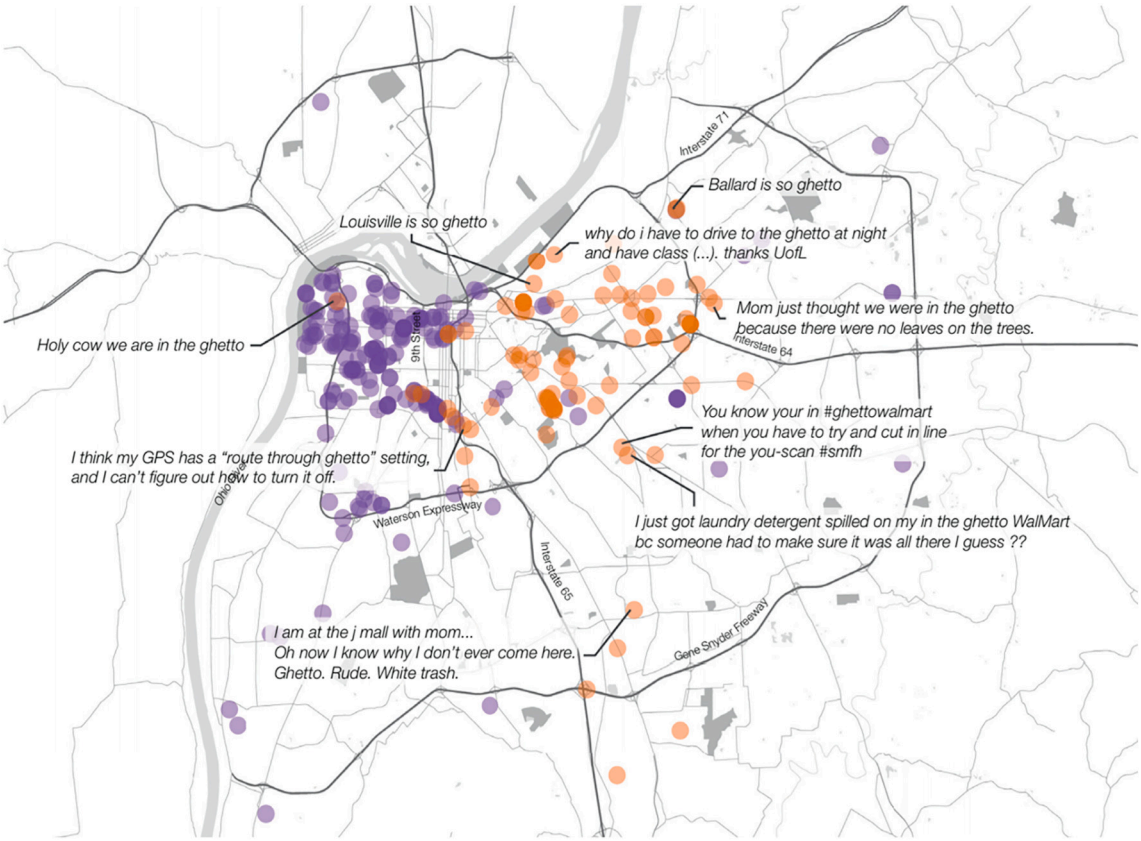

Fig. 6. The dots represent tweets from west-end (purple) and east-end (orange) Twitter users in Louisville. The addition of text to the tweets on the map enables the authors to explore where and why users cross the "9th Street divide." (by Shelton et al).

how people express themselves on each particular media (and thereby conduct meaningful qualitative studies), it is a good strategy for getting as many geotags as possible. In these studies, the social web offers a "big-N opportunity" rather than a phenomenological one.

Second, the situated boundaries papers similarly couch their problem frames in behavioral terms. However, their insistence on harnessing the spatial granularity of data changes the way behavioral problems are phrased and, consequently, the types of experts that it makes sense to invite in to help solve them. First, people's behavior is not assumed to be explained with reference to administrative zones. For instance, the Livehoods project offers alternative spatial divisions to the established administrative grid. A consequence of re-ontologizing the city in this way is that the relevant experts on urban life are no longer the decision makers associated with specific administrative units, such as Shadyside. Rather, the experts is the crowd from which the new urban boundaries are sourced. This is why the Livehoods project uses interviews with people on the street to understand the background of the clusters and their development over time. By showing this potential, the situated boundaries papers flip the script on who need to be involved in solving urban issues. Situated boundaries call for the engagement of local knowledge in ways that bureaucratic grids do not. Though the papers in the situated boundaries category do not take full advantage of the opportunity to include the voices of citizens directly from the traces they leave, they do open new debates about the spatial and administrative units within which to address a specific issue. Something that can also be done with standard sensors related to the ubiquitous computing of the smart city.

The phenomenological grid papers give rise to quite different problematizations than those discussed above. By framing urban space as a semantic network of meaning, the urban problem becomes one of sensemaking and lived experience rather than one of behavior. Notably, the character of this phenomenological re-problematization changes the higher these projects move up the metaphysical axis. This is clear in the comparison between two papers in the phenomenological grid that both explore the urban issue of safety. Quercia et al. (2015) used the thin measure of posting activity to create a quantitative and binary distinction between safe and unsafe streets, whereas Hu et al. (2019) used thick traces to make a more semantically enriched argument about what safety means to different people in different neighborhoods. The map in Fig. 4 fully utilizes the semantic thickness of review texts to map relations between topics, sentiments, and review scores associated with different spaces.

Such thick semantic projects escape the tendency to frame urban issues—like safety— as well-defined, binary, and quantified problems that can be measured as a number (creating a narrow space for urban solutions to be found in top-down urban planning, policy, or management initiatives that reduce or increase said number). Learning from design theory's distinction between well-defined and ill-defined 'wicked' problems (Buchanan, 1992), we might suggest that projects in the phenomenological grid differentiate themselves by not assuming an a priori understanding of urban issues. Instead, thick semantic projects like Hu et al.'s (2019) study are inspirational in demonstrating that social media data can be used to both qualify what an issue like safety means to different people and quantify patterns of how it is experienced. If we agree that issues such as safety can be construed as well-defined problems, then maybe this is not an issue. But if we believe that many of the important urban issues of tomorrow are open-ended, complex, and wicked in nature, formulating more qualified problems will also lead to more qualified urban governance. Such alternative problem formulations is the promise of projects in the phenomenological grid category. Something that ethnographic fieldwork can also do on a smaller scale.

The choice to stay low on the horizontal dimension and aggregate data within preexisting and standardized spatial geo-units makes it possible for the phenomenological grid papers to compare their insights to established urban issues. This process creates what Espeland and Stevens (1998) calls a situation of *commensurability*. It enables semantic and qualitative data to exist in the same ontological world as more recognizable demographic urban data. For instance, Hu et al. (2019: 1052) explicitly perform what they call "correlation analyses between the subjective perceptions extracted from [social media] and the objective socioeconomic attributes of New York City neighborhoods." They deliberately use geographical commensurability to make their semantic insights actionable for city planners, politicians, and developers. The paper feeds on a genre of cadastral mapping that has, for decades, served as a shared reference between different forms of urban expertise. However, sticking to the established grid also causes a reproduction of modes of categorizing the city without much curiosity about how digital traces might offer alternative issue-specific ways of drawing city boundaries. While these papers could have used the granularity of their semantic data to create situated boundaries of safety, this would also risk the commensurability that makes these maps understandable and useful for the stakeholders in urban planning. In that sense, the phenomenological grid shift the problem framing, but not the experts. The decision-makers and bureaucrats in established administrative units will still have a strong agency in proposing solutions.

The strategic upside of commensurability is lost in SCS papers that move out on both dimensions simultaneously. As noted in the previous section, these studies tend to deconstruct rather than enrich existing cartographies. Whereas the other three approaches in their own ways have synergies with established urban analyses, SCS papers carve out their own agenda. In doing so, they demonstrate the potential of fully utilizing the two affordances of social media data identified in this article. Namely, how it opens for re-drawing of issue-specific urban boundaries from the bottom up, while empowering researchers and planners to map the city as a phenomenologically layered multitude of lived experiences. While this arguably paves the way for innovative cartographies, it also raises the risk of being too alien for practitioners of urban planning. If SCS projects fail to translate data into insights that are actionable for actors in the field, they can become so explorative that they unsettle administrative operationality. In the face of decision making, complexity and ambiguity are enemies, whether in the form of re-drawn spatial units that transcend established grids or soft phenomenological mappings of user subjectivities without any hook to standard modes of analyzing. As many scholars aim for their results to have an impact, it is understandable that few papers in our review move all the way out on the two dimensions in Fig. 2 and use social media data to break completely with established standards in urban analysis. This reveals that the ambition of an SCS research agenda is difficult to pursue and realize as long as the use of social media data and geospatial mapping exercises are evaluated against standards inherited from established data paradigms. Paving the way for an impactful SCS agenda therefore necessitates revisiting some of the epistemological commitments that currently serve as standards for how to "properly" do data projects. We will end this paper by doing just that.

6. Revisiting the epistemological pillars of data-driven urbanism

In the previous section, we showed that an expansion of the types of data used in data-driven urbanism has the potential to re-problematize the city in interesting ways. But our proposed SCS agenda comes with the need to rethink some of the epistemological criteria used to evaluate the quality of urban data sources and algorithmic techniques. Well-known criteria from statistics, such as the need for representative samples, validity in the operationalization, and replicability of the study design, are ill-suited to evaluate the knowledge produced in SCS. The relevance of these criteria has been discussed by practitioners of qualitative methods for decades. To construct a new epistemological ground for an SCS research agenda there is a need to learn from this debate. Furthermore, we are also well aware that recent literature has presented compelling arguments for the problems associated with using algorithmic techniques on big digital datasets (see e.g. Boyd and Crawford, 2012 or D'Ignazio and Klein, 2020) and provided interesting reflections on what we can even learn about urban life from digital traces (Schwartz & Halegoua, 2015). We need to learn from these debates as well. Finally, there is a need to take inspiration from the pioneering SCS works reviewed above as well as the way scholars in computational humanities and human geography have recently revisited central epistemological debates.

6.1. Rethinking bias

We have already seen that the issue of bias have been prevalent in critical discussions of Big Data and many of the studies in our review also notes that one of the main drawbacks of online data is the bias of the web (Bocconi et al., 2015; Hochman & Manovich, 2013). The fact that each platform has its own distinct user group is seen as a problem in building valid and reliable maps and models. Rather than being interested in how data from the social web reflect situated urban experiences, such data is evaluated on whether it is comprehensive and representative. In fact, we saw above how the sensor surrogate papers deliberately sourced data from many platforms in an attempt to alleviate this problem in the form of a triangulation move.

However, other studies in our review illustrate the potential for revisiting this stance on bias. For instance, in their attempt to define digital neighborhoods, Anselin and Williams (2016) actively utilized the bias of data from Twitter and Foursquare to operationalize digital hot spots as "those areas in the city where one is more likely to use social media" (Anselin and Williams, 2016: 6). The existence of a digital divide among users is used to make a larger claim about the digitality of New York neighborhoods and to identify digital deserts in need of attention. Rather than approaching Twitter and Foursquare as sampling techniques, they are treated as cases with well-known (biased) characteristics. It is the knowledge of these characteristics—not the representativity of the data—that enables the authors to reach their conclusions. This epistemological strategy of leveraging the inevitable bias of online data has also long been seen as a necessary methodological move within computational humanities (Bruns & Burgess, 2011; Madsen, 2012).

However, the discussion of bias does not just concern data sources. Recent literature has discussed how the training and use of algorithms determine how they discriminate between things in the world (Gillespie, 2014). If algorithms are trained on a biased subset of data, they will fail to recognize things that fall outside the norm in that specific dataset. The obvious reaction would be to think of this as a problem, but once again, this form of bias can be turned into a strength. Zhang et al. (2019) takes such an approach in their project, where they use online images to train algorithms to recognize visual characteristics of cities. By focusing their research on areas of cities where the algorithm fails, they turn the idea of predictive analytics on its head. The failure to predict indicates that the algorithm has stumbled upon interesting exceptions to the prevailing visual norm of the city in question. Again, this epistemological strategy of leveraging algorithmic bias is also in active use in the field of digital methods (Munk, Olesen, & Jacomy, 2022).

Our suggestion is not to blindly use digital data and algorithms without considering their biases. We find recent studies of racial and gendered bias in machine learning immensely important in the context in which they are put forward (for illustrative cases on bias in Amazons recruitment systems and contemporary image analysis see Crawford, 2021). But what we can learn from the literature on computational humanities, and the works of scholars like Anselin, Williams and Zhang, is that within an SCS framework it is also possible to recast bias as an analytic opportunity. As an imperative for getting closer to our data rather than a reason to undermine and cast aside results. Bias can tell us something interesting about the state of the systems that provide data if

we maintain a critical stance. In a future SCS research trajectory, it is thus crucial to replace the inherited tendency to see social media data as an unreliable source that equals bad representativity just because it does not necessarily include everyone in a population. The point is to utilize the fact that each digital trace is produced by a human with relevant subjective experiences, feelings, emotions, and perceptions and to be creative about how we can use techniques from machine learning in an explorative way.

6.2. Rethinking the empirical ground

Another strong epistemological stance that needs to be revisited is the idea that the offline automatically trumps the online as an epistemological ground. As a ghost from the days of cyberspace, it is still custom to think of the web as something distinct from—more artificial and fake—than the physical world (Rogers, 2013). This assumption leads to an epistemological strategy of always constructing offline baselines from which to judge the validity of online data. We see tendencies of this strategy in the phenomenological grid, where papers aim to cross-reference the phenomenological aspects of social media data with other city data. The assumption is that subjective experience- and perception-based social media data are not reliable or interesting in themselves, and that their validity and analytical significance emerges and increases from being cross-referenced with "hard" data on which statistical tests of representativity can be performed. This assumption probably also explains why open government data is the third largest data source in our reviewed papers (see Fig. 1). It meets these conditions.

This stance presents a hierarchy in which data from the social web can, at best, be supplements or surrogates for more valid data obtained from the physical world. However, studies on the emergence of digital cities (Halegoua, 2020; Schwartz & Halegoua, 2015) have effectively shown how software has joined the built environment at the center stage of urban life. Drawing on De Certeau and Randall (1984), we suggest interpreting digital acts, such as check-ins, collaborative ratings, and algorithmic recommendations, as tactics that digital citizens use to weave space together in new ways and re-embed themselves in a fragmented and fluctuating urban landscape. The city is hybrid—it crosses the digital and the physical in a way that makes this distinction fruitless as an epistemological quality criteria. To pave the way for an SCS research agenda, we might benefit from rethinking the idea that traces on social media are mirrors of life in the analog world and instead accept that social media constitutes important arenas for our urban lifeworld. This approach to digital traces echoes Lev Manovich's (2020) argument that today's social media not only reflects the analogue cultural experience we have in concerts and when reading books but also makes up a huge part of these cultural experiences.

6.3. Rethinking the map

A third important epistemological discussion concerns the ontological status of maps in urban planning. A distinction can be made between those who strive to produce maps that represent objective realities and those who aim to make the subjective layers of the city visible. The usual epistemological stance on the practice of cartography would suggest the former, but work in critical human geography has reminded us that maps are not 1:1 representations of the world, nor are they ever neutral or objective. For instance, Krygier and Wood (2011) claimed that the work of mapmakers involves proposing different—often conflicting —realitiesof the world, highlighting some details while omitting others in the process of reducing the complexity of the world to make it navigable. Similarly, Kitchin and Dodge (2007) propose to think of maps as practices rather than products: "Maps are constantly in a state of becoming; they are ontogentic (emergent) in nature" (Kitchin and Dodge, 2007, 340).

We strongly agree with this assessment, even though not everyone does. In fact, there is a differentiating parameter in our reviewed corpus of literature, where many papers present data-driven maps as seemingly objective, neutral, and factual representations of the world, while others assign less ontological security to the maps they produce and present them as propositions that are off-the-moment, transitory and context-dependent outcomes of exploring social media traces. In papers by Weiler et al. (2013) and Xia et al. (2014), where the purpose is to provide users—specifically journalists—with real-time event detection and information about the city's ongoing activities, the cartography of social media traces is never fully formed and the work of producing maps is never complete.

Once again, this approach is mirrored in contemporary work within computational humanities, where data visualizations and maps have explicitly been produced with the aim to disturb—rather than solidify—existing knowledge claims (Madsen & Munk, 2019; Munk et al., 2019). We suggest replacing the aspiration to produce objective maps of the urban world with the ambition of proposing maps as off-the-moment representations of people's lived experiences. In so doing, the quality criteria by which to evaluate the validity of maps are no longer just their factual truthfulness but also the degree to which the propositions they make about lived experience can be recognized by the people whose life worlds they supposedly map. This echoes recent calls to involve people in the making of metrics and maps that concern them (Jensen et al., 2021; Williams, 2020; D'Ignazio & Klein, 2020) and even allow people to talk back to maps when they disagree with them (Niederer, 2018).

7. Conclusion

This paper started out by arguing that data-driven urbanism has become entangled with the smart city and thereby practiced in a way that prioritizes control over physical objects and downplays the human, semantic and political aspects of data. We labeled this approach to data-driven urbanism 'hard city sensing' and we drew on existing critical scholarship to note its shortcomings. However, we also argued that the emergence of the 'digital city' alters the empirical potentials for data-driven urbanism in a way that makes it possible to untangle it from the HCS framework. In making this argument we took inspiration from the way scholars within the computational humanities have utilized the existence of digital traces to produce data-driven research with a phenomenological and explorative outset. We noted that scholars in this tradition have primarily taken advantage of two affordances of digital traces. One is the fact that such traces are simultaneously qualitative and quantitative. This means that algorithms can be used to find patterns in semantic aspects of the world. The other is the fact that such traces are granular enough to work from an explorative epistemology and source classifications from data rather than pre-defined schemes.

We used these affordances as the backbone of a literature review of two decades of scholarship that have pioneered the use of digital traces as an empirical foundation for producing urban insights and cartographies. We see these studies as indications of the way new data possibilities are being grasped within the field of urban studies. By plotting each paper on two affordance-dimensions, we identified four ways of working with digital traces in an urban context. The 'sensor surrogates' approach analyzes digital traces in a framework that reminds of hard city sensing. The 'phenomenological grid' approach takes advantage of the qualitative affordance of digital traces, but not their explorative potential. The 'situated boundaries' approach takes advantage of the explorative affordance, but not the qualitative one. The SCS approach take advantage of both affordances simultaneously.

After having identified these approaches, we discussed the extent to which each of them paves the ground for new ways of problematizing the city and stimulates a need for inviting new experts into data-driven urban planning. While we argue that three of the approaches in their own distinct ways carry such potentials, we also noted that their future impact requires urban scholars and decision makers to revisit established epistemological stances on bias in data, the legitimacy of different

forms of empirical grounds and the ontology of urban cartographies. Instead of abandoning the idea of data-driven urbanism, we propose 'Soft City Sensing' as a framework for re-engaging with it with inspiration from the computational humanities and the pioneering studies in our review. On the one hand we urge qualitative urban scholars to ponder the possibilities of furthering their urban interest by beginning to 'think with algorithms' while retaining their qualitative and interpretative ambitions. Data-driven urbanism needs inputs from qualitative researchers just as much as such researchers can expand their toolkit with new methods. Even though the involved data and techniques have serious problems when it comes to bias we propose that such problems can be turned into analytical potentials if the urban analyst remains curious about the dynamics of bias and explores such biases to reflect on the societies and tools that produced them. On the other hand, we urge urban decision makers to expand their criteria for what serves as valid data inputs to urban planning. Whereas explorative SCS maps cannot serve the same functions as smart city dashboards, they are valuable tools in the important phase where urban problems are defined. In this phase maps and data visualizations are not produced to provide evidence and answers, but rather as effective tools with which frame meaningful questions about the urban realm and ensure that subsequent interventions are suited to the context in which they are supposed to make a positive change.

CRediT authorship contribution statement

Anders Koed Madsen: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Anders Grundtvig:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Sofie Thorsen:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors state that there have been no conflicting interests in writing this paper and all three authors agree on the division of authorship.

References

Alexander, C. (1968). A city is not a tree. *Ekistics, 139*, 344–348.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics, 214*(1), 481–518.

Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine, 12*.

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.

Bruns, A., & Burgess, J. (2011). The use of Twitter hashtags in the formation of ad hoc publics. In *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011* (p. 1). The European Consortium for Political Research (ECPR).

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues, 8*(2), 5–21.

Caprotti, F. (2019). Spaces of visibility in the smart city: Flagship urban spaces and the smart urban imaginary. *Urban Studies, 56*(12), 2465–2479.

Cardullo, P., & Kitchin, R. (2019). Smart urbanism and smart citizenship: The neoliberal logic of 'citizen-focused' smart cities in Europe. *Environment and Planning C: Politics and Space, 37*(5), 813–830.

Congress for the New Urbanism. (2000). Charter of the new urbanism. *Bulletin of Science, Technology & Society, 20*(4), 339–341.

Crawford, K. (2021). *The atlas of AI*. Yale University Press.

Cukier, K., Mayer-Schönberger, V., & de Véricourt, F. (2021). *Framers: Human advantage in an age of technology and turmoil*. Penguin.

De Certeau, M., & Randall, S. (1984). Walking in the city. In R. Guins, & O. Cruz (Eds.), *Popular culture: A reader* (pp. 449–461). SAGE.

Dewey, R. (1950). The neighborhood, urban ecology, and city planners. *American Sociological Review, 15*(4), 502–507.

D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology, 24*(1), 313–343.

Galster, G. C. (1986). What is neighbourhood? An externality-space approach. *International Journal of Urban and Regional Research, 10*(2), 243–263.

Gehl, J. (2011). *Life between buildings: Using public space*. Island press.

Gillespie, T. (2014). The relevance of algorithms. *Media Technologies: Essays on Communication, Materiality, and Society, 167*(2014), 167.

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

Goodwin, C. (2015). Professional vision. In *Aufmerksamkeit* (pp. 387–425). Springer.

Halegoua, G. R. (2020). *The digital city: Media and the social production of place*. NYU Press.

Harding, A., & Blokland, T. (2014). *Urban theory: A critical introduction to power, cities and urbanism in the 21st century*. Sage.

Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice, 2*(1), 47–62.

Jensen, T. E., Birkbak, A., Madsen, A. K., & Munk, A. K. (2021). Participatory data design: Acting in a digital world. In T. Zuiderent-Jerak, & G. Downey (Eds.), *Making and doing STS*. MIT Press.

Kitchin, R. (2014a). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

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

Kitchin, R. (2017). Data-driven urbanism. In *Data and the city* (pp. 44–56). Routledge.

Kitchin, R., & Dodge, M. (2007). Rethinking maps. *Progress in Human Geography, 31*(3), 331–344.

Kitchin, R. (2015). Making sense of smart cities: addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society, 8*(1), 131–136. https://doi.org/10.1093/cjres/rsu027

Krygier, J., & Wood, D. (2011). Ce n'est pas le monde (This is not the world). In M. Dodge, R. Kitchin, & C. Perkins (Eds.), *Rethinking maps: New frontiers in cartographic theory*. Routledge.

Kryssanov, V. V., Okabe, M., Kakusho, K., & Minoh, M. (2001). Communication of social agents and the digital city—A semiotic perspective. In *, 56. Kyoto Workshop on Digital Cities*. Springer.

Kurgan, L. (2013). In *Close up at a distance: Mapping, technology, and politics.*. MIT Press.

Kurgan, L., Brawley, D., et al. (2019). In *Ways of Knowing Cities*. Columbia Books on Architecture and the City.

Lynch, K. (1960). *The image of the city*. MIT press.

Madsen, A. K. (2012). Web-visions as controversy-lenses. *Interdisciplinary Science Reviews, 37*(1), 51–68.

Madsen, A. K., & Munk, A. K. (2019). Experiments with a data-public: Moving digital methods into critical proximity with political practice. *Big Data & Society, 6*(1), Article 2053951718825357.

Manovich, L. (2020). *Cultural analytics*. MIT Press.

Marvin, S., & Luque-Ayala, A. (2017). Urban operating systems: Diagramming the city. *International Journal of Urban and Regional Research, 41*(1), 84–103.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Moretti, F. (2013). *Distant reading*. Verso Books.

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs.

Munk, A. K., & Jensen, T. E. (2015). Revisiting the histories of mapping. *Ethnologia Europaea, 44*(2), 31.

Munk, A. K., Madsen, A. K., & Jacomy, M. (2019). Thinking through the databody. In Å. Mäkitalo, T. Nicewonger, & M. Elam (Eds.), *Designs for experimentation and inquiry: Approaching learning and knowing in digital transformation* (p. 110). Routledge.

Munk, A. K., Olesen, A. G., & Jacomy, M. (2022). The Thick Machine: Anthropological AI between explanation and explication. *Big Data & Society, 9*(1) (in review).

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research, 49*(1), 3–42.

Niederer, S. (2018). *Networked images: Visual methodologies for the digital age*.

Raban, J. (1974). *Soft City: What Cities Do To Us, and How They Change the Way We Live, Think and Feel.*. London: Hamish Hamilton.

Rogers, R. (2013). *Digital methods*. MIT press.

Schwartz, R., & Halegoua, G. R. (2015). The spatial self: Location-based identity performance on social media. *New Media & Society, 17*(10), 1643–1660.

Scott, J. C. (2020). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale university press.

Thorsen, S., & Munk, A. K. (2022). *Beyond the geotag: Opening new empirical routes in urban studies with visual social media and digital methods to chart human-space relations* (Forthcoming paper).

Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company.

Vanolo, A. (2014). Smartmentality: The smart city as disciplinary strategy. *Urban Studies, 51*(5), 883–898.

Venturini, T., & Latour, B. (2010). The social fabric: Digital footprints and qualiquantitative methods. *Proceedings of Futur en Seine*, 87–101.

de Waal, M. (2017). A city is not a galaxy: Understanding the city through urban data. In *Data and the city* (pp. 17–30). Routledge.

Whyte, W. H. (1980). *The social life of small urban spaces*. Washington, DC: Conservation Foundation.

Williams, S. (2020). *Data action: Using data for public good*. MIT Press.

Zuiderent-Jerak, T. (2015). *Situated intervention: Sociological experiments in health care*. MIT Press.

Reviewed papers

Adelfio, M., Serrano-Estrada, L., Martí-Ciriquián, P., Kain, J. H., & Stenberg, J. (2020). Social activity in Gothenburg's intermediate city: Mapping third places through social media data. *Applied Spatial Analysis and Policy, 13*(4), 985–1017.

Alfayez, A., Awwad, Z., Kerr, C., Alrashed, N., Williams, S., & Al-Wabil, A. (2017, July). Understanding gendered spaces using social media data. In *International Conference on Social Computing and Social Media* (pp. 338–356). Cham: Springer.

Anselin, L., & Williams, S. (September 30, 2015). Digital neighborhoods. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 9*(4), 305–328.

Bocconi, S., Bozzon, A., Psyllidis, A., Titos Bolivar, C., & Houben, G. J. (2015, May). Social glass: A platform for urban analytics and decision-making through heterogeneous social data. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 175–178). ACM.

Boy, J. D., & Uitermark, J. (2017). Reassembling the city through Instagram. *Transactions of the Institute of British Geographers, 42*(4), 612–624.

Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012, May). The livehoods project: Utilizing social media to understand the dynamics of a city. In *Sixth International AAAI Conference on Weblogs and Social Media*.

Currid, E., & Williams, S. (2009). The geography of buzz: Art, culture and the social milieu in Los Angeles and New York. *Journal of Economic Geography, 10*(3), 423–451.

De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., & Lepri, B. (2016, April). The death and life of great Italian cities: A mobile phone data perspective. In *Proceedings of the 25th International conference on World Wide Web* (pp. 413–423). International World Wide Web Conferences Steering Committee.

Del Bimbo, A., Ferracani, A., Pezzatini, D., D'Amato, F., & Sereni, M. (2014, April). Livecities: Revealing the pulse of cities by location-based social networks venues and users analysis. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 163–166). ACM.

García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gomez, B., Condeco-Melhorado, A., & Gutierrrez, J. *The pulse of the city through Twitter: Relationships between land use and spatiotemporal demographics*. (2018). arXiv preprint arXiv:1705.07956.

García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gomez, B., Condeco-Melhorado, A., Gutierrez, J., et al. (2018). City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities, 72*, 310–319.

Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday, 18*(7).

Hu, Y., Deng, C., & Zhou, Z. (2019). A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers, 109*(4), 1052–1073.

Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a collective sense of place. *PLoS One, 11*(4), Article e0152932.

Jones, P., Layard, A., Speed, C., & Lorne, C. (2015). MapLocal: Use of smartphones for crowdsourced planning. *Planning Practice and Research, 30*(3), 322–336.

Kelley, M. J. (2013). The emergent urban imaginaries of geosocial media. *GeoJournal, 78*(1), 181–203.

Kim, H. J., Chae, B. K., & Park, S. B. (2018). Exploring public space through social media: An exploratory case study on the High Line New York City. *Urban Design International, 23*(2), 69–85.

Kloeckl, K., Senn, O., & Ratti, C. (2012). Enabling the real-time city: LIVE Singapore! *Journal of Urban Technology, 19*(2), 89–112.

Kozaki, Y., Wang, Y., & Kawai, Y. (2018, October). Generating pictorial maps for tourists using Flickr photo data. In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)* (pp. 403–407). IEEE.

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

Li, Y., & Shan, J. (2013). Understanding the spatio-temporal pattern of tweets. *Photogrammetric Engineering and Remote Sensing, 79*, 769–773.

Longley, P. A., & Adnan, M. (2016). Geo-temporal Twitter demographics. *International Journal of Geographical Information Science, 30*(2), 369–389.

Madsen, A. K. (2022). *Do you live in a bubble? Designing critical diversity metrics to intervene in urban cartography, Projections, MIT Press*.

Manovich, L., Tifentale, A., Yazdani, M., & Chow, J. (2014). The exceptional and the everyday: 144 hours in Kyiv. In *The 2nd Workshop on Big Humanities Data held in conjunction with IEEE Big Data 2014 Conference* (Paper in PDF format) http://www.the-everyday.net/.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q., & Vespignani, A. (2013). The twitter of babel: Mapping world languages through microblogging platforms. *PLoS One, 8*(4), e61981.

Netto, V. M., Pinheiro, M., Meirelles, J. V., & Leite, H. (2015). Digital footprints in the cityscape: Finding networks of segregation through Big Data. In *International Conference on Location-based Social Media Data* (pp. 1–15).

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS One, 7*(5), e37027.

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

Peng, X., Bao, Y., & Huang, Z. (2020). Perceiving Beijing's "city image" across different groups based on geotagged social media data. *IEEE Access, 8*, 93868–93881.

Peng, Z., Wang, R., Liu, L., & Wu, H. (2020). Exploring urban spatial features of COVID-19 transmission in Wuhan based on social media data. *ISPRS International Journal of Geo-Information, 9*(6), 402.

Poorthuis, A., & Zook, M. (2014). Artists and bankers and hipsters, oh my! Mapping tweets in the New York Metropolitan Region. *Cityscape, 16*(2), 169–172.

Psyllidis, A., Bozzon, A., Bocconi, S., & Bolivar, C. T. (2015, July). A platform for urban analytics and semantic data integration in city planning. In *International Conference on Computer-aided Architectural Design Futures* (pp. 21–36). Berlin, Heidelberg: Springer.

Quercia, D., Aiello, L. M., Schifanella, R., & Davies, A. (2015, May). The digital life of walkable streets. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 875–884). International World Wide Web Conferences Steering Committee.

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

Rossi, L., Boscaro, E., & Torsello, A. (2018, April). Venice through the lens of Instagram: A visual narrative of tourism in Venice. In *Companion Proceedings of the The Web Conference 2018* (pp. 1190–1197).

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.

Smarzaro, R., Lima, T. F. D. M., & Davis, C. A., Jr. (2017, April). Could data from location-based social networks be used to support urban planning?. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 1463–1468). International World Wide Web Conferences Steering Committee.

Sobolevsky, S., Bojic, I., Belyi, A., Sitko, I., Hawelka, B., Arias, J. M., & Ratti, C. (2015, June). Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity. In *2015 IEEE International Congress on Big Data* (pp. 600–607). IEEE.

Song, X. P., Richards, D. R., He, P., & Tan, P. Y. (2020). Does geo-located social media reflect the visit frequency of urban parks? A city-wide analysis using the count and content of photographs. *Landscape and Urban Planning, 203*, 103908.

Song, Y., Fernandez, J., & Wang, T. (2020). Understanding perceived site qualities and experiences of urban public spaces: A case study of social media reviews in Bryant park, New York city. *Sustainability, 12*(19), 8036.

Tifentale, A., & Manovich, L. (2015). Selfiecity: Exploring photography and self-fashioning in social media. In *Postdigital aesthetics* (pp. 109–122). London: Palgrave Macmillan.

Weiler, A., Scholl, M. H., Wanner, F., & Rohrdantz, C. (2013, June). Event identification for local areas using social media streaming data. In *Proceedings of the ACM SIGMOD Workshop on Databases and Social Networks* (pp. 1–6). ACM.

Williams, S. (2015). *Here now! Social media and the psychological city*.

Williams, S., Xu, W., Tan, S. B., Foster, M. J., & Chen, C. (2019). Ghost cities of China: Identifying urban vacancy through social media data. *Cities, 94*, 275–285.

Xia, C., Schwartz, R., Xie, K., Krebs, A., Langdon, A., Ting, J., & Naaman, M. (2014, April). CityBeat: Real-time social media visualization of hyper-local city data. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 167–170). ACM.

Zhang, F., Zhou, B., Ratti, C., & Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. *Royal Society Open Science, 6*(3), 181375.