



RESEARCH ARTICLE



Mapping dynamic human sentiments of heat exposure with location-based social media data

Fangzheng Lyu^{a,b} , Lixuanwu Zhou^{a,b}, Jinwoo Park^c, Furqan Baig^{a,b} and Shaowen Wang^{a,b}

^aCyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Urbana, IL, USA; ^bDepartment of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA; ^cDepartment of Geography and Geographic Information Science, University of North Dakota, Grand Forks, ND, USA

ABSTRACT

Understanding urban heat exposure dynamics is critical for public health, urban management, and climate change resilience. Near real-time analysis of urban heat enables quick decision-making and timely resource allocation, thereby enhancing the well-being of urban residents, especially during heatwaves or electricity shortages. To serve this purpose, we develop a cyberGIS framework to analyze and visualize human sentiments of heat exposure dynamically based on near real-time location-based social media (LBSM) data. Large volumes and low-cost LBSM data, together with a content analysis algorithm based on natural language processing are used effectively to generate near real-time heat exposure maps from human sentiments on social media at both city and national scales with km spatial resolution and census tract spatial unit. We conducted a case study to visualize and analyze human sentiments of heat exposure in Chicago and the United States in September 2021. Enabled with high-performance computing, dynamic visualization of heat exposure is achieved with fine spatiotemporal scales while heat exposure detected from social media data can be used to understand heat exposure from a human perspective and allow timely responses to extreme heat.

HIGHLIGHTS

- Near real-time and high spatial resolution mapping of human sentiments of heat exposure with Twitter data
- An integrated cyberGIS and machine learning framework for visualizing heat exposure with Twitter data
- Human sentiment of heat exposure mapping in the City of Chicago and the United States

ARTICLE HISTORY

Received 25 December 2022
Accepted 10 April 2024

KEYWORDS

CyberGIS; heat exposure; location-based social media; urban heat

1. Introduction

Heat-related illnesses, ranging from mild heat edema, heat syncope to severe heat-stroke, have become a major contributing factor to a significant increase in severe diseases and death (Goodman *et al.* 2001, Lugo-Amador *et al.* 2004, Thorsson *et al.* 2014, Shindell *et al.* 2020). These conditions are especially alarming in urban areas where temperature is typically higher than surrounding areas due to Urban Heat Island (UHI) effects. According to the United States Census Bureau, more than 80% of the country's population resides in urban areas and thus is directly influenced by UHI effects. Evaluating and mapping extreme heat exposure is therefore vital for ensuring the health and well-being of urban populations.

The proliferation of ubiquitous low-cost location technologies has provided new opportunities for a variety of volunteered geographic information (VGI) research studies to analyze spatial and spatial-temporal aspects of human populations (Goodchild 2007). In particular, Location-Based Social Media (LBSM) data has become a versatile data source for solving urban problems (Sui and Goodchild 2011, Martí *et al.* 2019, McKittrick *et al.* 2022, Niu and Silva 2023) due to its wide availability. Combined with advanced natural language processing (NLP) techniques, LBSM data can be analyzed to extract the extent of hotness expressed by its content. Different from the conventional metrics such as land surface temperature for the quantification of heat exposure, unique insights into the individual experience of heat can be gained by examining social media posts. This alternative approach offers a distinct perspective, emphasizing the human dimension of heat exposure. Moreover, the analysis of the 'human perspective' regarding heat exposure, enabled by near real-time data, can facilitate the near real-time identification of heat-susceptible areas. However, despite its usefulness and value, LBSM exhibits unique characteristics and poses challenges associated with processing and managing a large volume and variety of data.

In this context, a key research question is 'how to evaluate and map human sentiments of heat exposure?' To answer this question, this study aims to explore the possibility of using LBSM data to map human sentiments of heat exposure by analyzing Twitter data. The study offers three major advantages. Firstly, by enabling real-time retrieval of LBSM data coupled with fast analysis enabled by high-performance computing, the study enables near real-time capabilities for the mapping and assessment of human sentiments of heat exposure. Such dynamic evaluation of urban heat can be critical for timely decision-making, especially in response to extreme heat events. Secondly, this approach addresses urban heat from a human-centric standpoint, recognizing the variability in individuals' heat tolerance depending on their location. Despite attempts to integrate different factors, accurately evaluating people's sentiments of heat remains challenging. In this study, social media data offers a promising tool for effectively capturing human sentiments of heat exposure. Lastly, Twitter data, being both low-cost and widely available, allows for the extension of near real-time mapping of heat exposure from urban to national scales. Overall, we argue that near real-time heat exposure mapping from social media data can serve as an effective approach to understand heat exposure from a human perspective. Furthermore, analysis of dynamic social media data can support timely responses to extreme heat and optimize resource allocations in areas with the most urgent needs.

With the support by cyberGIS which is defined as cyber-based geographic information science and systems (cyberGIS) (Wang 2010, Wang *et al.* 2021), we have designed a framework to visualize and analyze human sentiments of heat exposure both in the City of Chicago and the United States with Twitter data collected in real-time. Dynamic visualization and analysis of city-level heat exposure were shown with different temporal granularity. Depending on the needs of users, this framework can be generalized for detecting human sentiments of heat exposure in different spatiotemporal contexts, representing scales and granularity with different input parameters. To ensure reproducibility, this near real-time analysis is conducted with CyberGIS-Jupyter, which takes advantage of high-performance computing resources while providing a consistent scientific workflow environment (Wang 2010, Yin *et al.* 2019, Lyu *et al.* 2021). The following Section 2 reviews related literature. Section 3 describes the pipeline for Twitter data collection and explains the semantic analysis method we employed to understand the ‘heat index’ of the content for each Tweet message. We provide an example of mapping human sentiments of heat exposure in near real-time for both Chicago and the US at different spatial resolutions and evaluating heat exposure results compared with temperature and the Social Vulnerability Index in Section 4. Section 5 presents a discussion of the proposed human sentiments of heat exposure. Finally, the reproducibility of this work is enabled with the support of cyberGIS.

2. Related work

The measurement of heat exposure entails a multifaceted challenge that encompasses both human and environmental factors. From the environmental side, the mapping of urban heat distribution is straightforward with temperature data. The National Oceanic and Atmospheric Administration (NOAA) provides a graphical forecast of temperature across the United States every three hours (<https://graphical.weather.gov/sectors/southplains.php>). However, the limited number of weather stations in urban environments results in insufficient spatial granularity in the heat information derived from collected temperature data (Wang *et al.* 2021). From the human side, with a focus on the adverse health impacts of extreme heat and to provide decision support with resource allocation in preparation for and response to heat-related events, social vulnerability is widely used to indicate the potential negative effects on communities caused by external stresses on *human* health, such as extreme heat (Lehnert *et al.* 2020). The social vulnerability index (SVI) and heat vulnerability index (HVI) have been extensively studied to identify the socially vulnerable population exposed to extreme heat at the county level across the US (Reid *et al.* 2009, Cutter *et al.* 2003, Conlon *et al.* 2020, Niu *et al.* 2021). However, since both the SVI and HVI are derived from yearly releases of US census data, they do not offer real-time or near real-time information. In our study, we aim to generate high spatial resolution and near real-time heat exposure maps using people’s expressions about heat on social media.

With 221 million daily active users in the third quarter of 2021 (Twitter 2021), Twitter data have been widely used for scientific research for its low cost, large volume, and wide availability (McKittrick *et al.* 2022). Human thoughts and emotions presented in social media data like Twitter data are widely studied in the domain of NLP

as sentiment analysis, where social media posts are assigned a ‘sentiment value’ based on the interpreted meaning of the contents (Martin and Schuurman 2017, Karmegam and Mappillairaju 2020, Zivanovic *et al.* 2020, Camacho *et al.* 2021, Kovács *et al.* 2021, Omuya *et al.* 2023). However, Twitter data are often questioned in scientific research for its bias and representativeness in terms of (1) geographical distribution, where populous counties are overrepresented with Twitter users (Mislove *et al.* 2021); (2) age distribution, where only 17.1 percent of Twitter users are above the age of 50 (Statistica 2021a); (3) gender, where 70.4 percent of the Twitter users are male (Statistica 2021b); and (4) race, such as under-sampling of Hispanic users in the south-west; under-sampling of African-American users in the south and Midwest; and the oversampling of Caucasian users in major cities (Mislove *et al.* 2021). It is estimated that up to 15% of Twitter users were automated bot accounts (Varol *et al.* 2017). Despite the limitations, Twitter data has been a vital resource for geographers and environmental scientists thanks to its extremely large volume and rich geospatial information. Given that about three percent of all tweets are geotagged with native location information available (Leetaru *et al.* 2013, Sloan and Morgan 2015, Twitter 2020), the data resource has been applied to understand heatwave (Cecinati *et al.* 2019, Wang *et al.* 2021, Zander *et al.* 2023), natural hazards (Bee *et al.* 2012, Slamet *et al.* 2018, Vishwanath *et al.* 2023), snow depth (Muller 2013), flood (Fohringer *et al.* 2015, Li *et al.* 2018, Li *et al.* 2023), influenza (Gao *et al.* 2018), etc. In the urban context, previous research demonstrated that the large-scale and near real-time observations are unique advantages of LBSM data and they are effective in better understanding urban sustainability in general (Ilieva and McPhearson 2018). Specifically, LBSM data was used to assess different urban dynamics (Martí *et al.* 2019), such as human mobility (Luo *et al.* 2016), place of interest identification (Van Canneyt *et al.* 2012), crowd behavior (Lee *et al.* 2013), event detection (Béjar *et al.* 2016), and decision support for city design (Dunkel 2015).

At the same time, real-time analysis of social media is made possible with GIS. Sui and Goodchild (2011) suggest that a more dynamic process of time-critical or real-time monitoring and dynamic decision-making is possible with the fusion of GIS and LBSM data. In the context of real-time detection and visualization using GIS, Xia *et al.* (2014) visualize and detect spatial trends for events from geotagged photos within New York City in a real-time fashion and Lyu *et al.* (2022) detect near real-time UHI with urban sensing networks in the City of Chicago.

A cyberGIS framework, which integrates advanced cyberinfrastructure, geographic information systems, and spatial analysis, is adopted in this study to achieve scalable, reproducible, and high spatial resolution and near real-time analysis of human sentiments of heat (Wang 2010). CyberGIS has been widely adopted in data- and computation-intensive geospatial problem-solving in domains such as digital twins (Shirowzhan *et al.* 2020), public health (Padmanabhan *et al.* 2013), hydrology (Liu *et al.* 2018, Lyu *et al.* 2019, 2021), and disaster management (Han *et al.* 2023). In this study, we employ CyberGIS-Jupyter, a platform that leverages the capabilities of ROGER, the first geospatial supercomputer, to facilitate data-intensive, reproducible, and scalable geospatial analytics within a Jupyter Notebook environment (Wang 2016, Yin *et al.* 2019). In particular, by involving high-performance computing and a keyword-based natural

language processing approach, this study enables real-time collection and evaluation of extensive social media data for assessing heat exposure.

3. Data and method

There are four major categories for retrieving and understanding human sentiments in the domain of NLP: (1) keyword spotting classifies text based on the presence of unambiguous affect words; (2) lexical affinity involves a probable ‘affinity’ for each word; (3) statistical methods include Bayesian inference and support vector machines in semantic analysis; and (4) concept-based approaches use web ontologies or semantic networks to understand contents (Cambria *et al.* 2013). While the computational demands of these approaches limit their real-time implementation, keyword spotting stands out due to its accessibility and computational efficiency, particularly for real-time and near real-time analysis (Kumar and Subba 2020). Thus, a keyword-based approach is utilized in this study to understand Twitter posts with a heat dictionary generated with Word2vec, which is a well-known and extensively used tool in keyword-based sentimental analysis (Xue *et al.* 2014, Al-Saqqah and Awajan 2019).

The scientific workflow of this study is shown in Figure 1. First, the heat dictionary comprises a compilation of weather-related words and phrases paired with corresponding values ranging from −1 to 1, indicating their relevance to the notion of ‘hot weather’. The greater the value, the more it is relevant to ‘hot weather.’ Real-time LBSM data collection, heat dictionary, and mapping of human sentiments of heat exposure are incorporated within the framework to realize near real-time heat exposure visualization at both city and national scales.

3.1. Study area

This study focuses on the City of Chicago as the designated area for city-level analysis. The UHI phenomenon in the Chicago region has been extensively investigated in prior research. Chen *et al.* (2022) provide estimates of heat-related exposure and assess the impact of UHI during 2012 Chicago heatwave. Lyu *et al.* (2021) focus on predicting

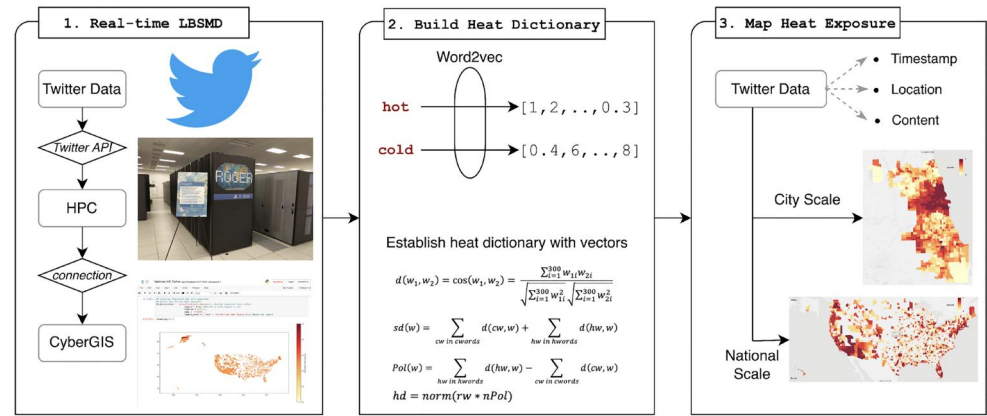


Figure 1. Scientific workflow.

UHI in Chicago with a fine spatiotemporal resolution. Alfraihat *et al.* (2016) evaluate UHI in Chicago from an ecological perspective. Additionally, Johnson *et al.* (2012) utilize a heat vulnerability index to identify vulnerable groups in Chicago with regard to extreme heat.

Despite the mitigating influence of the lake breeze from Lake Michigan, it is predicted that the future heatwaves in Chicago during the latter half of the 21st century will be characterized by increased intensity, frequency, and duration (Meehl and Tebaldi 2004, Sharma *et al.* 2016). Furthermore, the analysis at the national level is conducted across the United States, employing counties as the spatial unit.

3.2. Data

Out of privacy concerns, a scientific workflow was built using CyberGIS-Jupyter to allow indirect usage of Twitter data. A data collection pipeline is set up on ROGER, the first geospatial supercomputer (Wang 2016), to collect and store large-scale Twitter data in real-time (See [Supplemental Material A](#) for technical details). In addition to Twitter data, the One Billion Words Benchmark for Language Modeling is used for the establishment of a dictionary with word2vec (Chelba and Mikolov 2014, Xue *et al.* 2014). National-level shapefile for the US and census-tract level shapefile for the City of Chicago are collected from the Chicago Data Portal and United States Census Bureau respectively for analyzing and visualizing heat exposure. Temperature data from weather stations in Chicago are collected from the National Center for Environmental Information in the National Oceanic Atmospheric Administration (NOAA) and weather underground (<https://www.wunderground.com/>).

3.3. Establish heat dictionary

The first step in the process is to establish a heat dictionary for words relevant to heat exposure. The dictionary is then used to analyze the textual content of the Twitter data. We use the word2vec tool, which has been widely used in the domain of sentimental analysis and urban functions region extraction (Xue *et al.* 2014, Zhai *et al.* 2018). Using word2vec, we generated 300 dimensions of vector representation for about 3 million words and automatically derived phrases after 9 hours of training on ROGER (See [Supplemental Material B](#) for technical details).

In order to evaluate the heat index of Tweet content, we established a word dictionary for heat from vector representation (Xue *et al.* 2014). First, we defined a set of synonymic words for hot weather as *hwords* and another set of synonymic words for cold weather as *cwords*. Initially, we compiled a comprehensive list of synonyms for 'hot weather' and 'cold weather' by referencing Dictionary.com, which is primarily based on the Random House Unabridged Dictionary. Subsequently, we consulted the Oxford dictionary to generate the sets of words pertaining to hot weather (*hwords*) and cold weather (*cwords*). Based on the dictionary result, *hwords* included warm, balmy, summery, tropical, boiling, blazing, baking, scorching, roasting, searing, flaming, parching, blistering, sweltering, torrid, sultry, humid, muggy, airless, oppressive, stifling while *cwords*, on the other hand, included chilly, cold, cool, crisp, fresh, brisk, bleak,

wintry, snowy, frosty, icy, glacial, polar, arctic, raw, sharp, bitter, biting, piercing, penetrating, freezing, frigid. Furthermore, we also defined the similarity distance (d) between two words (w_1 and w_2) as their cosine value (Equation (3)).

$$hwords = \{A \text{ set of words for hot weather}\} \quad (1)$$

$$cwords = \{A \text{ set of words for cold weather}\} \quad (2)$$

$$d(w_1, w_2) = \cos(w_1, w_2) = \frac{\sum_{i=1}^{300} w_{1i} w_{2i}}{\sqrt{\sum_{i=1}^{300} w_{1i}^2} \sqrt{\sum_{i=1}^{300} w_{2i}^2}} \quad (3)$$

Using the similarity, we distinguished the relatedness of the words in Twitter data to cold or hot weather by finding the summation of similarity distance (sd) from these words to words from $cwords$ and words from $hwords$. More formally, the summation of similarity distance (sd) is defined as (Equation (4)):

$$sd(w) = \sum_{cw \text{ in } cwords} d(cw, w) + \sum_{hw \text{ in } hwords} d(hw, w) \quad (4)$$

A given word is more weather-related if the value of the summation of its similarity distance is greater. We further calculate rw as the normalized summation of similarity distance for all words in the word bank by normalizing their value from 0 to 1 with 1 being most related to weather and 0 being least related. And rw , which represents the word's relatedness to weather, is defined as (Equation (5)):

$$rw = \text{norm}(sd) \quad (5)$$

As words like 'the' and 'I' do not describe weather, we further rule out words that are distantly away from weather by subsetting the largest 20,000 words/phrases out of the 300 million words/phrases in the entire word bank. In addition, for those 20,000 words/phrases that are weather-related, we further distinguish their polarization, which is defined as Pol in the equation that follows, towards hot weather by subtracting the similarity distance to all $cwords$ to $hwords$:

$$Pol(w) = \sum_{hw \text{ in } hwords} d(hw, w) - \sum_{cw \text{ in } cwords} d(cw, w) \quad (6)$$

Then, we define $nPol$ as the normalized polarization representing the word's polarization toward hot weather. $nPol$ is a normalized value between -1 to 1 where -1 represents polarization toward cold weather and 1 is polarization toward hot weather (Equation (7)):

$$nPol = \text{norm}(Pol) \quad (7)$$

Finally, we define the heat dictionary (hd) as a normalized value between -1 and 1 of the products of the relatedness of every word in the bank to weather (rw) and its polarization toward hot and cold weather ($nPol$). The greater the normalized value of a word in the dictionary, the more it is relevant to 'heat':

$$hd = \text{norm}(rw * nPol) \quad (8)$$

3.4. Content analysis

The raw Twitter data, collected using the Twitter Application Programming Interface (API), is stored in JSON text format. Each Tweet post encompasses over a hundred fields, incorporating essential and nested attributes. These fields encompass details such as timestamp, content, location, number of likes, username, and more. Specifically, location information is delineated by either an exact point represented through a pair of coordinates or a multi-polygon comprising sets of coordinate pairs as nodes (Twitter n.d.). For the purpose of this study, when collecting data on ROGER, we filtered out tweets without geotags and used relevant fields including timestamp, geospatial location, and textual content only. We computed heat exposure (hd) for every tweet post by adding the heat index of all the words in the textual content using the heat dictionary described in the previous section:

$$he = \sum_{w \text{ in } \text{twitter_content}} hd(w) \quad (9)$$

3.5. Spatial mapping

Once we have heat exposure estimated for each Tweet post, we use the spatial and temporal information from tweets to map the near real-time data at the city as well as the national scale. In both national-level and city-level analyses, a filtering process is employed to select Twitter posts based on their geospatial location and temporal information. The heat exposure value is then calculated by aggregating the heat exposure values obtained from averaging all weather-related tweets within each spatial unit. For city-level analysis, the spatial units are defined as 1 km x 1 km grids and census tracts, while national-level analysis utilizes counties as the spatial unit.

For city-scale analysis, we were able to derive a high spatial resolution heat exposure map since cities generally tend to have larger populations and thus more geotagged tweets at different locations within a short period of time. Location information in Twitter data can be retrieved in two forms: (1) points, which consist of coordinates (i.e. longitude and latitude) representing the geographic location of a specific tweet as posted by a user, and (2) multi-polygons, which represent the associated place mentioned in a tweet. Multi-polygons are typically defined by four pairs of coordinates, delineating the geographical extent encompassing the location scope. In the latter case, it does not entail specific location information, so we ran Monte-Carlo simulations to distribute the location randomly within the bounding box of the multi-polygon to reduce the bias (Mooney 1997). The analysis could be potentially more accurate with more simulations. However, in this research, we choose 100 simulations to balance the accuracy and limitation of computational power. For every simulation, as all the Twitter posts were pinpointed to a certain location, we used the inverse-distance weighting (IDW) method to transform discrete spatial heat exposure points from tweets into a spatial field. Finally, we mapped the near real-time human sentiments of heat exposure by averaging the values from all simulations. Depending on the number of simulations we conducted, this process could be computationally intensive. Powerful computational resources and advanced parallel

computing techniques are desirable as we want to generate high-accuracy maps without significant delay in time.

Contrary to city-scale analysis, we chose to have a coarser spatial resolution for national-scale mapping of human sentiments of heat exposure. With available data, even at this coarse spatial resolution there existed sparsely populated counties having limited social media data recorded for real-time analysis.

4. Results

4.1. City-scale heat exposure map

One city-scale analysis was conducted using data collected on September 27th, 2021, in the City of Chicago. Unexpected hot weather was detected in Chicago with the highest temperature being 88 degrees Fahrenheit based on data from NOAA. We performed this near real-time analysis at 12 p.m. Central Time (CT). There are in total 21,525 geo-referenced tweets collected during a 48-h span. Among these 21,525 geotagged tweets, 772 of them are determined as weather-related, each of which has at least one word matching another word in the heat dictionary. 26 tweets among 772 are recorded with an exact location while the other 746 come with a multi-polygon as their georeferenced coordinates. To ensure a comprehensive representation of the study area using the LBSM data, there is a tradeoff between the quantity of available data and the selected spatial granularity in the proposed model. The proposed model can achieve finer spatial granularity with additional data collection within the proposed timeframe. In this case study, both census tracts and 1 km spatial resolution, with approximately 800 spatial units each for the Cook County, are selected to ensure comprehensive area representation with the amount of social media getting collected. [Figure 2](#) maps the human sentiments of heat exposure with quantile classification generated using 1 km as spatial resolution and census tract as a spatial unit. As shown in [Figure 2a](#), the center for heat exposure is located just north of downtown Chicago. The southern part of the city is generally reporting low-level heat exposure at that time. A high-resolution map can be generated as it provides valuable information about the city to help avoid extreme heat especially in the summertime or during the heatwave.

Furthermore, we conducted a similar analysis at the census tract level. As we can see in [Figure 2b](#), our study enables dynamic detection of heat exposure for each census tract area in the City of Chicago based on real-time collected georeferenced Twitter data. Out of 801 census tracts in Chicago, census tracts 3102 and 3103 located on the lower west side of Chicago and census tracts 5501 and 5502 located in the southeast corner of Chicago, are showing significant low heat exposure. Further demographic investigation shows that the lower west side of Chicago has a large group of Hispanic population and there are many Mexican restaurants and bars on the side of South Halsted Street in that area. Similarly, more than half of the population living in the southeast corner of Chicago are Hispanic. Based on ACS data in 2020, in Census tracts 3102, 3103, 5501 and 5502, 43%, 66%, 55%, and 58% of the population are Hispanic while the average is 29% in Chicago. From the perspective of language, 26%, 63%, 42%, and 36% of the adult population in those census tracts are Spanish-speaking respectively, while the average is 22% in the City of Chicago. Not limited to the

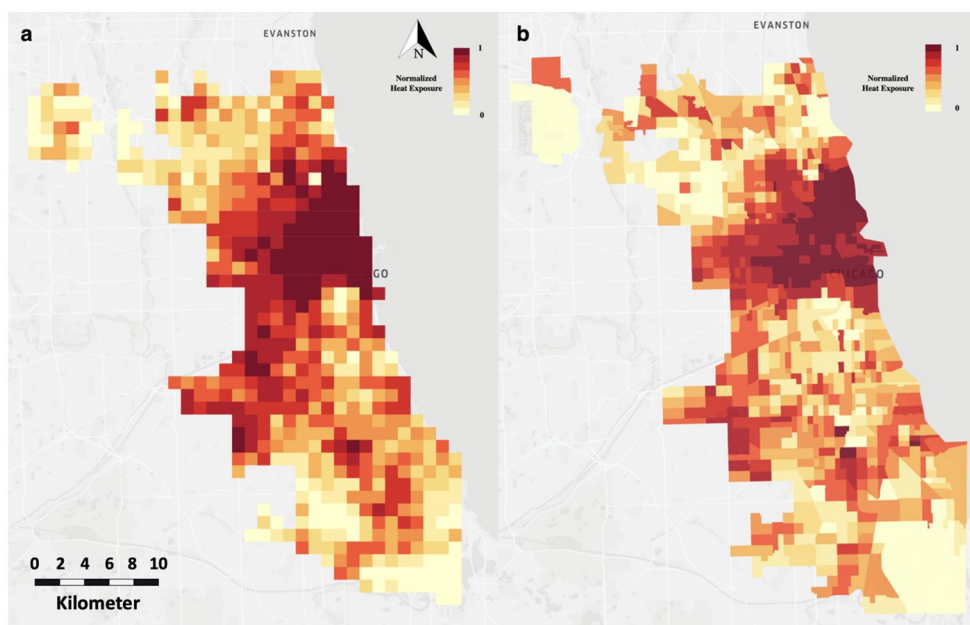


Figure 2. Heat exposure map for the City of Chicago on 2021.9.27: (a) 1 km*1km spatial resolution and (b) census tract as the spatial unit.

City of Chicago, such a heat exposure map can be efficiently generated given the spatial scope of another city. Analysis and visualization of heat exposure can potentially help better understand heat exposure from a human perspective while dynamic heat exposure analysis and visualization can help with timely resource allocation and decision-making to mitigate losses caused by extreme heat.

4.2. Dynamic changes in heat exposure

Dynamic monitoring of human sentiments of heat exposure from Twitter data was conducted from September 21st to September 27th to detect how people responded on Twitter to temperature rises and falls as well as the spatial distribution of extreme hot and cold areas. We analyzed the Twitter contents at 12 p.m. (CT) on each day taking into consideration all the Tweets posted within 24 h range. In particular, from September 22nd to Sept 27th, approximately 350 heat-related social media data points were consistently gathered within 24 h in the study area. After normalizing all the data collected on those days, we use Jenks nature breaks to generate [Figure 3](#).

From a temporal perspective, the heat exposure was increasing as temperature gradually increased from September 23rd to September 27th. A significant heat cluster is spotted on September 25th. It was not the hottest day but there was a sudden increase in temperature happening during the afternoon of September 24th. On the other hand, on September 22nd, the heat exposure in Chicago was the lowest, which was caused by a temperature decrease from September 21st to September 22nd. As a result, not only the extreme hot and cold weather triggered the responses on Twitter data, but the sudden increase and decrease of temperature could be a deciding factor

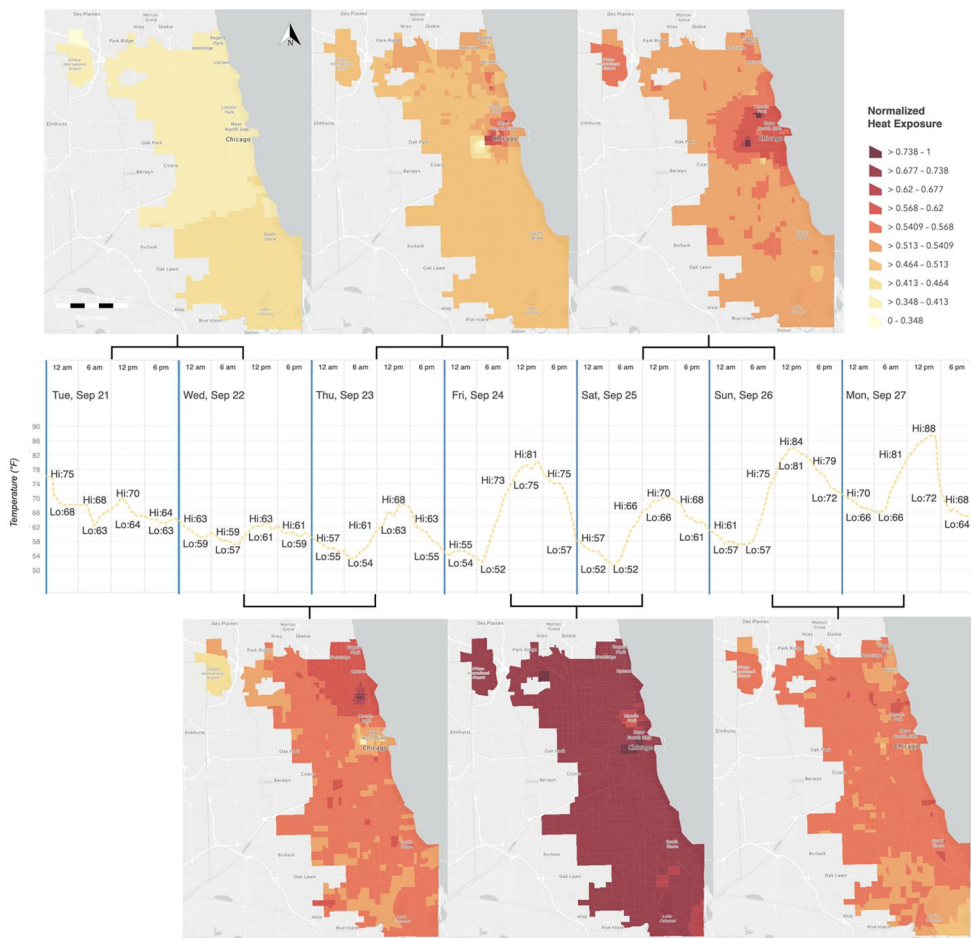


Figure 3. Normalized Heat Exposure from September 21, 2021 to September 27, 2021.

as well. The dynamic monitoring of heat exposure from Twitter, compared with regular temperature, is more sensitive than temperature changes.

From a spatial perspective, on September 24th, 25th, and 26th, there was a clear heat cluster located near the downtown area of Chicago. This is consistent with the findings by Lyu *et al.* (2021) indicating the UHI effect gets worse in and around downtown Chicago. We demonstrated that during heat waves or sudden temperature increases, the downtown area should get more attention.

Additionally, the Twitter-based heat exposure dynamics can be analyzed at a finer temporal resolution. In this study, we further looked at how heat exposure changes on September 26th, 2021. The reason we are particularly interested in this day is the significant temperature rises in Chicago where the highest temperature increased from 70 degrees Fahrenheit to 84 degrees Fahrenheit from September 25th, 2021, to September 26th, 2021. We investigated the dynamics of heat exposure at a fine temporal resolution to gain insight into the changes in people's exposure to dramatic temperature rises within a day. As shown in Figure 4, the human sentiments of heat exposure for these separate time periods are plotted using 1km*1km cell as spatial

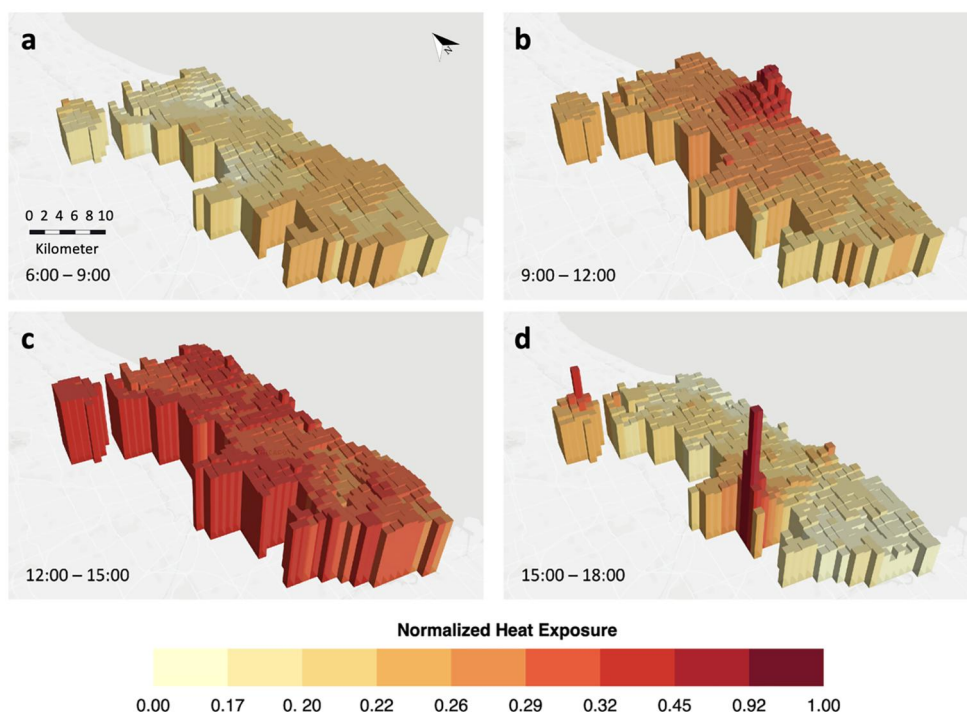


Figure 4. Normalized Twitter-based heat exposure on September 26th, 2021, in the City of Chicago from 9:00 to 18:00 with 1km*1km cell as the size of spatial units, (a) 6:00–9:00; (b) 9:00–12:00; (c) 12:00–15:00; (d) 15:00–18:00.

units during daytime from 6:00 to 18:00 when the temperature was dramatically increasing and fell after 14:00. A 3-h time interval is selected to make sure enough data were collected to represent the entire Chicago area. From 6:00–9:00 (Figure 4a), the overall heat exposure was relatively low and the area with higher heat exposure was mostly located in the suburban area on the south side. From 9:00–12:00 (Figure 4b), the temperature gradually increased, and there were areas reporting relatively high heat exposure near the downtown area of Chicago. From 12:00–15:00 (Figure 4c), the temperature reached a maximum of 81 degrees Fahrenheit. During this period, the heat exposure increased across the city and it was the hottest period during that day. Lastly, from 15:00–18:00 (Figure 4d), the heat exposure dropped with decreasing temperature while two clear heat clusters were spotted at the O'Hare International Airport and Midway International Airport.

Overall, dynamic, near real-time, and high spatiotemporal resolution monitoring of heat exposure in the city can be achieved to detect and analyze heat exposure based on Twitter data.

4.3. Comparison between human sentiments of heat, temperature and SVI

Human sentiments of heat exposure capture subjective perceptions of temperature and related environmental conditions derived from human-generated content within

social media data. As such, this value encompasses both environmental and human dimensions. In this section, a comparison is conducted between human sentiments of heat extracted from Twitter data and environmental factors (based on the spatial mapping of heat exposure derived from temperature data) as well as human factors (based on SVI). The aim is to address the following questions: (1) Statistically, what are the similarities and differences between human sentiments of heat and SVI? (2) Does the distribution of heat exposure obtained from Twitter data align with the patterns observed from temperature and SVI data?

To answer these two questions, the minimum, maximum and average temperature of 9.26 and 9.27 for each census tract area is interpolated using an Inverse Distance Weighted method based on in total of 59 functional NOAA weather stations and underground weather stations and sensors within the City of Chicago. Among all functional weather stations, 8 of them are from NOAA and 51 of them are from the underground weather sensors network. [Figure 5](#) shows the relationship between normalized heat exposure from LBSM data and the quantile map of minimum, maximum and average temperature on 9.26 and 9.27 in Chicago. For all 6 temperature indicators, there are positive correlation between the indicator and heat exposure showing the areas with high heat exposure reported from LBSM data are mostly places with relatively higher temperatures. The spatial distribution of temperature indicators is mostly consistent with the spatial distribution of normalized heat exposure from Twitter data, where the 'heat' area is centered around the north of downtown Chicago.

[Table 1](#) shows the Pearson and Spearman correlation between normalized heat exposure and the minimum, maximum and average temperature on 9.26 and 9.27. All the temperature indicators show a strong positive correlation with normalized heat exposure as the coefficients are positive and the p -value is less than 0.005. It is evident that the temperature is highly consistent with the value of heat exposure. Considering the multitude of factors influencing the perceived 'heat' among urban populations and the non-linear nature of the relationship between temperature and heat exposure, our analysis reveals a strong spatial and statistical correlation between human sentiments of heat exposure and temperature.

The correlation between the Social Vulnerability Index (SVI) and human sentiments of heat exposure during the period from September 22, 2021 to September 27, 2021 is shown in [Figure 6](#) and [Table 2](#). The SVI data utilized is sourced from the Centers for Disease Control and Prevention (CDC) for the year 2020. [Figure 6](#) illustrates a nearly horizontal line representing the correlation between SVI and human sentiments of heat exposure, indicating a weak correlation between the two variables. [Table 2](#) shows that although the p -value for the correlation is significant, the Pearson and Spearman coefficients have small absolute values. Consequently, unlike temperature, there is no strong correlation between the SVI and human sentiments of heat exposure.

Despite the overall weak correlation, [Figure 6](#) highlights an interesting observation on September 22, 2021. On this day, with a decrease in temperature and widespread complaints about cold weather in the city, a positive correlation is observed between the SVI and human sentiments of heat exposure. However, from September 23 to September 27, as temperatures was rising, the correlation between human sentiments of heat exposure and SVI is mostly negative, except for September 25. This

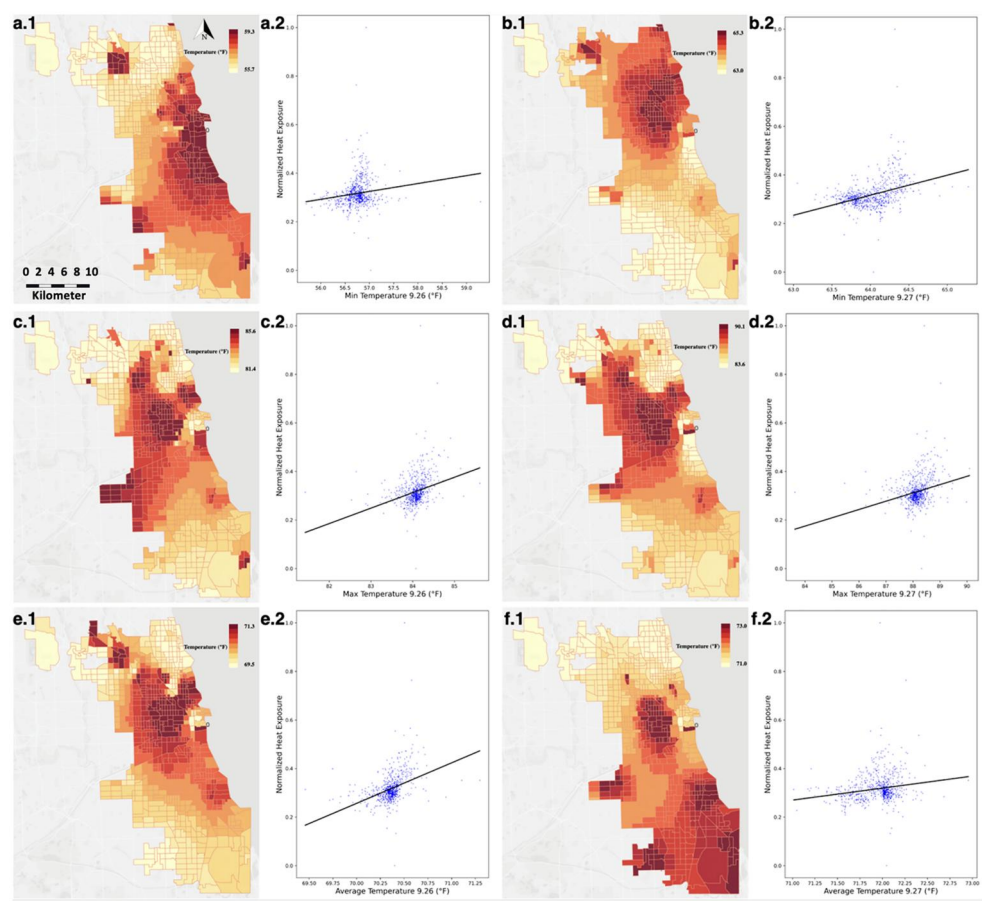


Figure 5. The relationships between normalized heat exposure and minimum, maximum and average temperature measured with weather stations on 9.26 and 9.27: (a.1) minimum temperature on 9.26, (a.2) the relation between minimum temperature on 9.26 and normalized heat exposure; (b.1) minimum temperature on 9.27, (b.2) the relation between minimum temperature on 9.27 and normalized heat exposure; (c.1) maximum temperature on 9.26, (c.2) the relation between maximum temperature on 9.26 and normalized heat exposure; (d.1) maximum temperature on 9.27, (d.2) the relation between maximum temperature on 9.27 and normalized heat exposure; (e.1) average temperature on 9.26, (e.2) the relation between average temperature on 9.26 and normalized heat exposure; (f.1) average temperature on 9.27, (f.2) the relation between average temperature on 9.27 and normalized heat exposure.

Table 1. Correlation between normalized heat exposure and minimum, maximum and average temperature on 9.26 and 9.27.

	Pearson coefficient	Pearson <i>p</i> -value	Spearman coefficient	Spearman <i>p</i> -value
Minimum temperature on 9.26	0.1482	2.5496e-05	0.1885	7.6150e-08
Maximum temperature on 9.26	0.2948	1.6051e-17	0.3677	4.7670e-27
Average temperature on 9.26	0.3860	3.9708e-29	0.4506	2.6571e-41
Minimum temperature on 9.27	0.3814	4.6399e-13	0.3736	6.2276e-28
Maximum temperature on 9.27	0.2519	7.2599e-30	0.2802	6.4606e-16
Average temperature on 9.27	0.1844	1.4678e-07	0.1494	2.1895e-05

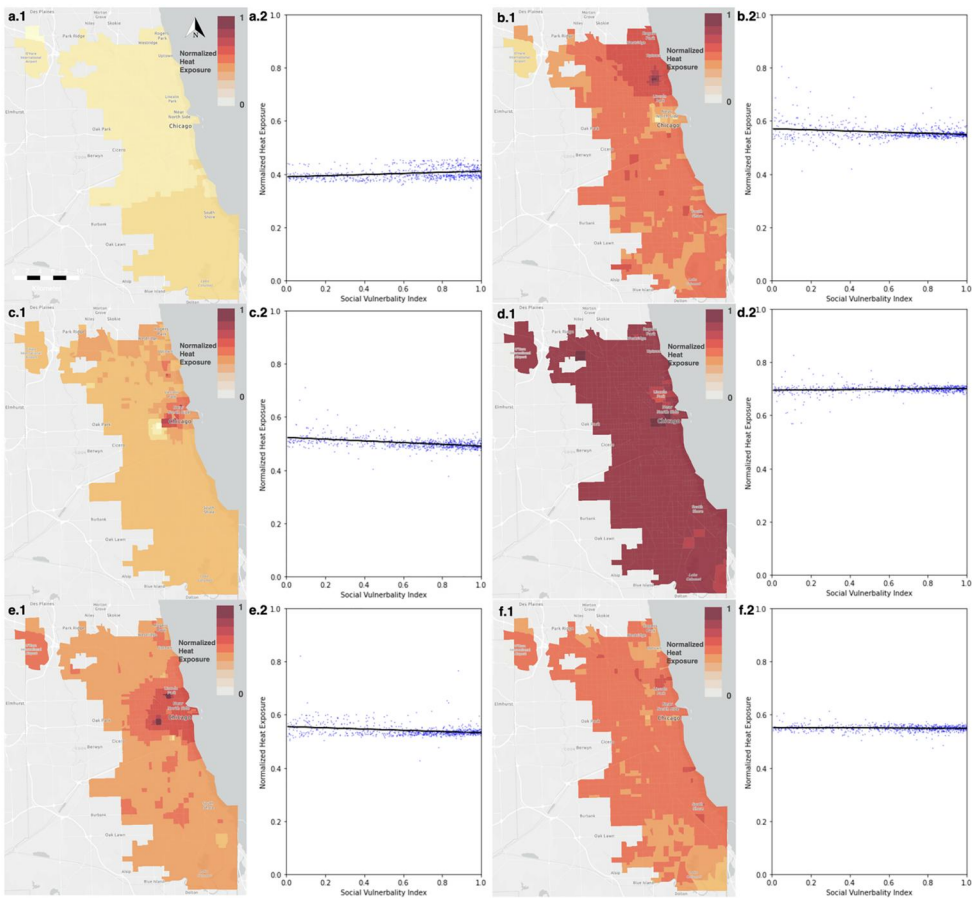


Figure 6. The relationships between SVI and normalized heat exposure from 2021.09.22 to 2021.09.27: (a.1) normalized human sentiments of heat exposure on 9.22, (a.2) the relation between normalized human sentiments of heat exposure on 9.22 and SVI; (b.1) normalized human sentiments of heat exposure on 9.23, (b.2) the relation between normalized human sentiments of heat exposure on 9.22 and SVI; (c.1) normalized human sentiments of heat exposure on 9.24, (c.2) the relation between normalized human sentiments of heat exposure on 9.24 and SVI; (d.1) normalized human sentiments of heat exposure on 9.25, (d.2) the relation between normalized human sentiments of heat exposure on 9.25 and SVI; (e.1) normalized human sentiments of heat exposure on 9.26, (e.2) the relation between normalized human sentiments of heat exposure on 9.26 and SVI; (f.1) normalized human sentiments of heat exposure on 9.27, (f.2) the relation between normalized human sentiments of heat exposure on 9.27 and SVI.

Table 2. Correlation between normalized heat exposure and SVI.

	Pearson coefficient	Pearson <i>p</i> -value	Spearman coefficient	Spearman <i>p</i> -value
2021.09.22	0.2796	1.6490e-15	0.2896	1.40234e-16
2021.09.23	−0.2263	1.5313e-10	−0.1113	1.829e-3
2021.09.24	−0.4336	3.9118e-37	−0.4400	2.551e-38
2021.09.25	0.1231	5.6345e-4	0.0915	1.050e-2
2021.09.26	−0.2826	8.0125e-16	−0.1356	1.4340e-4
2021.09.27	−0.0807	2.4027e-2	−0.0301	4.0012e-1

observation indicates that regions with higher Social Vulnerability Index (SVI) scores demonstrate a comparatively delayed responsiveness on social media when compared to less vulnerable regions, both in response to temperature increase and decrease.

4.4. Mapping national-scale heat exposure

Mapping was also conducted on September 27th, 2021, for the conterminous United States together with Hawaii and Alaska. Twitter data collected within the time ranging from 12 p.m. September 26th to 12 p.m. September 27th is taken into consideration. The extensive heat-related LBSM data is utilized to evaluate heat exposure at a national scale (See [Supplemental Materials C](#) for the amount of data collected for national-level analysis). The national-level heat exposure, as shown in [Figure 7](#), is calculated based on all the heat-related tweets collected within the 24-h period in each county. Based on our analysis, Franklin County in Maine is reported to have the lowest heat exposure, and Assumption Parish LA is reported to have the highest heat exposure. Counterintuitively, there is no clear trend that the southern part of the US, where temperature is generally higher, reported higher heat exposure than the northern part of the US. As people living in warm places are more used to hot weather, it is not likely for them to post on Twitter about hot weather if the high temperature is normal. More likely, people will post hot-related tweets to complain about extreme heat exposure when unexpected hot weather hits or there are sudden heatwaves or electricity shortages. [Figure 7](#) presents a national-level map of heat exposure for the Contiguous United States (CONUS). It is useful in terms of finding the places for sudden extreme heat fluctuation during a short period of time and those with significantly higher exposure compared with the surrounding regional areas. If we zoom in to examine a single state or counties that are adjacent to each other, where the

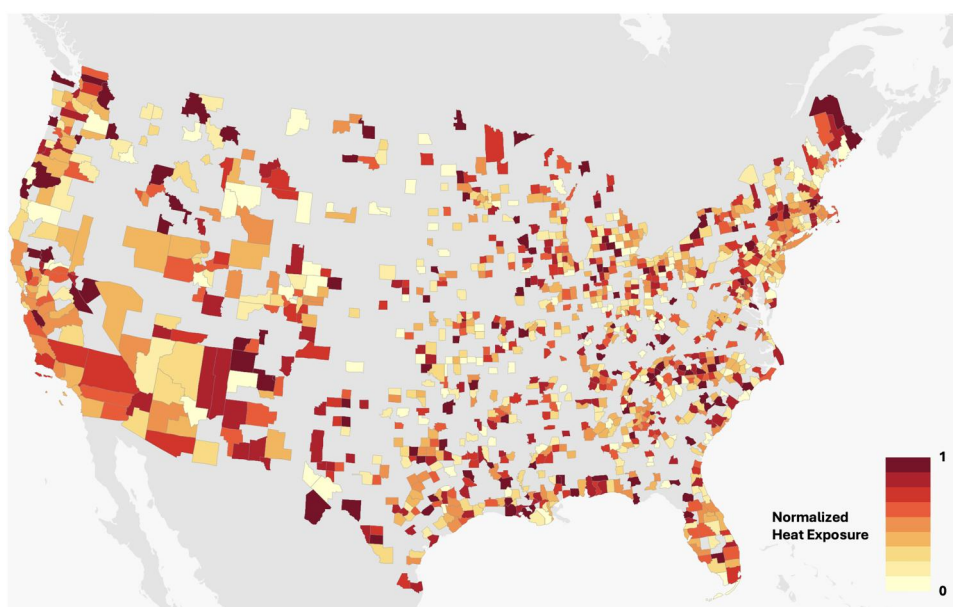


Figure 7. A national-level heat exposure map in the United States on 2021.9.27.

temperature is generally similar, we can find more detailed information about heat exposure in those localized areas. For example, Cook County where the City of Chicago is located, that day, has higher normalized heat exposure compared with the surrounding counties. That might be caused by the unexpected hot weather in Chicago that day with the highest temperature being 88 Degree Fahrenheit. More knowledge can be gained from long-term monitoring of heat exposure in a single county. Sudden dramatic unusual increase or decrease in heat exposure would require additional attention for heat-related problems. To address the spatial heterogeneity challenge, greater emphasis should be placed on studying human sentiments of heat exposure at the regional level.

5. Discussion

Heat exposure represents a multifaceted challenge encompassing various physical environmental attributes, such as temperature, humidity (Zhao *et al.* 2014, Zhang *et al.* 2023), air quality (Li *et al.* 2018), land use (Li *et al.* 2021), vegetation index (Chow *et al.* 2012), and socioeconomic factors, including population density, poverty, ethnicity, and others (Harlan *et al.* 2013, Maier *et al.* 2014, Bao *et al.* 2015). In addition to these factors, we contend that human sentiments of heat exposure, extracted from LBSM data, provide valuable insights into heat exposure from an individual human perspective. One major research motivation is to detect sudden changes in heat exposure in a near real-time fashion. Using human sentiments of heat exposure, we can quickly find the locations that are under the threat of extreme heat. Especially in cities like Chicago with a large population, dynamic monitoring of heat exposure can be achieved with fine spatiotemporal granularity using location-based social media data. Furthermore, our analysis compares human sentiments of heat exposure with temperature and the Social Vulnerability Index (SVI). The results indicate a strong correlation with temperature while showing a weaker correlation between SVI and the human sentiments. Despite the observed strong correlation between human sentiments of heat exposure and temperature, areas where high temperatures do not correspond to increased human sentiments of heat exposure, such as the southern part of the City of Chicago on September 27th, 2021, require further investigation. This aspect presents a particularly interesting opportunity for future systemic analytics. Regarding the limited correlation between SVI and human sentiments of heat exposure, it is noteworthy that regions with higher vulnerability exhibit a slower response to both temperature increases and decreases compared to less vulnerable regions.

It is worth noting that uncertainty exists in the analysis results mainly due to the usage of human-generated social media data. First, both the temporal scale and spatial scale matter. Even though our research framework can be applied in near real-time, fine temporal scale might result in inadequacy of data to cover a large area, especially when there is a need for fine spatial granularity. The dynamic changes of heat exposure could be biased if there are not enough heat-related Twitter data to cover the entire spatial domain given a fine temporal scale. As a result, the balance between spatial and temporal resolutions is a key to appropriately derive and interpret results. Second, the space-time context is another important factor. From the temporal perspective, when analyzing the daily/weekly changes in heat exposure, results would

be more useful if there is a sudden change in temperature because the framework is sensitive to significant temperature changes and frequent individual posts on social media during extreme weather such as heatwaves. When analyzing heat exposure within a day, the amount of Twitter data collected at a different time in a day could vary significantly depending on the time of collection. Based on our experiments, generally more data are available during afternoon and evening times compared with morning and midnight within a day. As a shortage of data could introduce uncertainty to analytical results, the time of data summarization plays a significant role. From the spatial perspective, city-level analysis of heat exposure is likely more robust when it is carried out in a populous city due to the positive correlation between population size and the number of tweets posted within a short time span.

While the new policy changes by Twitter in November 2022 no longer allow users to share physical location information without authentication (Twitter 2022), which could help filter out some noisy data, a major limitation of the biased nature of Twitter data contributes to potential uncertainty in content interpretation with the semantic analysis method and location information used. To resolve this bias, data from other sources including urban sensors and weather stations could be used to supplement the Twitter data for evaluating heat exposure in future research. A possible solution would be to put a location-aware weighted coefficient, which is calculated based on multiple socioeconomic and demographic factors, to each tweet post to address the bias and representative issue of the tweet post. Furthermore, an additional significant constraint stems from the keyword-based algorithm employed to estimate human sentiments of heat exposure. This approach exhibits limitations in the following aspects: (1) its inability to identify affect-negated words, such as 'isn't' and 'doesn't'; (2) its potential failure to comprehend that affect, which could be conveyed through underlying meanings within sentences; and (3) its incapacity to discern different contextual meanings for polysemous words (Cambria *et al.* 2013). As a result, our ongoing research aims to use more sophisticated natural language processing algorithms to take into consideration the context of tweets. Computer vision algorithms can also be applied to images posted by social media users to better analyze their heat exposure given that images are increasingly posted on social media. In addition, further research can be conducted by long-time monitoring of heat exposure to discover causal effects of heat exposure and its relationships with other health, environmental, socioeconomic, and demographic factors. Additionally, we plan to employ human expressions from LBSM data for evaluating attributes that are difficult to measure, such as disasters and pandemics. We also intend to investigate the time-sensitivity and uncertainty of different locations and various information categories using human expression.

6. Conclusion

This research uses location-based social media data collected from Twitter to map dynamic human sentiments of heat exposure both at city and national levels. A new cyberGIS framework is developed to analyze and visualize locations with extreme heat exposure based on Twitter data in different spatiotemporal contexts with various scales and granularities. First, Twitter data was collected in near real-time using a cyberGIS-based workflow. Second, we used word2vec to establish a dictionary for 'heat

value'. Then, for all the geotagged Twitter data collected, we analyze their contents based on the heat dictionary to find the heat index for each tweet. At last, we map human sentiments of heat exposure based on the Twitter data at the city level as well as the US national level with the spatial unit of census tract and county respectively. A case study was conducted in September 2021. From the city-level case study on Chicago, we find that the area just north of downtown Chicago experienced high-level heat exposure while the Hispanic and Spanish-speaking population, especially those who live in the southwest side of Chicago and southeast corner of Chicago, are under-represented by the Twitter data. In evaluating the proposed heat index in relation to temperature and SVI, we found a strong correlation between human sentiments of heat exposure and temperature, while a weaker correlation with SVI. Dynamic monitoring of heat exposure for a week and changes in heat exposure at different times within a day were analyzed to show that Twitter data can be used effectively to detect heat exposure variations at fine spatial and temporal scales. This framework can be generalized for other cities and regions across the globe where Twitter is being used.

Overall, our research represents the first attempt to analyze and visualize dynamic human sentiments of heat exposure based on location-based social media data collected in near real-time. Together with traditional temperature measurements, near real-time heat exposure information derived from social media data can be used to understand heat exposure from a human perspective and help with timely resource allocation to mitigate losses and negative impacts caused by extreme heat. Computational reproducibility and scalability are enabled by CyberGIS-Jupyter through integrating high-performance computing capabilities and an open-access scientific workflow environment.

Acknowledgements

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Our computational work used ROGER, which is a geospatial supercomputer supported by the CyberGIS Center for Advanced Digital and Spatial Studies and the School of Earth, Society, and Environment at the University of Illinois at Urbana-Champaign.

Authors' contribution

The authors confirm their contribution to the paper as follows: Fangzheng Lyu: conceptualization, methodology, analysis, writing, revision; Lixuanwu Zhou: methodology, analysis; Jinwoo Park: methodology, writing; Furqan Baig: data collection, writing; Shaowen Wang: conceptualization, methodology, writing, revision, supervision, project administration, and funding acquisition.

Disclosure statement

The authors declare no conflict of interest.

Funding

This work is supported in part by the National Science Foundation (NSF) under grant numbers: 1833225, 2112356, and 2118329.

Notes on contributors

Fangzheng Lyu is a research assistant in the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign (UIUC). He holds a Ph.D. and an M.S. in Geography from UIUC and a B.E. from the University of Hong Kong in Computing Engineering. His research interests focus on GIS, urban dynamics, remote sensing, cyberGIS, and big data analysis.

Lixuanwu Zhou is an M.S. student in the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign. He holds a B.S. from Zhejiang University in Mathematics.

Jinwoo Park is an Assistant Professor in the Department of Geography and Geographic Information Science at the University of North Dakota. He holds a Ph.D. in Geography from Texas A&M University, and an M.S. in Geography from Kyung Hee University. His research focuses on advancing spatial accessibility measurements of urban infrastructure by uncovering temporal dynamics and uncertainty in data-rich environments.

Furqan Baig is a Research Programmer in the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign. He holds a Ph.D. in Computer Science from Stony Brook University, and an M.S. in Computer Science at the Lahore University of Management Sciences. His research interests include parallel and distributed computing, spatio-temporal data management, and cyberGIS.

Shaowen Wang is a Professor of the Department of Geography and Geographic Information Science, and founding director of the CyberGIS Center for Advanced Digital and Spatial Studies at the University of Illinois at Urbana-Champaign. His research interests focus on geographic information science and systems (GIS), cyberGIS, and complex social and environmental problems.

ORCID

Fangzheng Lyu  <http://orcid.org/0000-0001-5180-0380>

Shaowen Wang  <http://orcid.org/0000-0001-5848-590X>

Data and codes availability statement

The data and codes that support the findings of this research are available at: https://github.com/cybergis/real_time_heat_exposure_with_LBSMD. Sample data containing Twitter data collected on 2021.09.25 and 2021.09.26 can be found at <https://doi.org/10.6084/m9.figshare.21780065>. For access to the full dataset or conducting near real-time analysis, please contact corresponding author as Twitter data cannot be made publicly available due to privacy protection.

References

Alfraihat, R., Mulugeta, G., and Gala, T., 2016. Ecological evaluation of urban heat island in Chicago City, USA. *Journal of atmospheric pollution*, 4 (1), 23–29.

- Al-Saqqa, S., and Awajan, A., 2019. The use of word2vec model in sentiment analysis: A survey. In *Proceedings of the 2019 international conference on artificial intelligence, robotics and control*, 39–43.
- Bao, J., Li, X., and Yu, C., 2015. The construction and validation of the heat vulnerability index, a review. *International Journal of Environmental Research and Public Health*, 12 (7), 7220–7234.
- Bee, E., et al., 2012. Hazard map: real time hazard mapping using social media. In: *AGI Geocommunity 12: Sharing the Power of Place*, 18–20 Sept 2012, Nottingham, UK.
- Béjar, J., et al., 2016. Discovery of spatio-temporal patterns from location-based social networks. *Journal of Experimental & Theoretical Artificial Intelligence*, 28 (1–2), 313–329.
- Camacho, K., et al., 2021. Sentiment mapping: Point pattern analysis of sentiment classified Twitter data. *Cartography and Geographic Information Science*, 48 (3), 241–257.
- Cambria, E., et al., 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28 (2), 15–21.
- Cecinati, F., et al., 2019. Mining social media to identify heat waves. *International Journal of Environmental Research and Public Health*, 16 (5), 762.
- Chelba, C., Mikolov, T., et al., 2014. One billion word benchmark for measuring progress in statistical language modeling.
- Chen, K., et al., 2022. Estimating heat-related exposures and urban heat island impacts: A case study for the 2012 Chicago heatwave. *GeoHealth*, 6 (1), e2021GH000535.
- Chow, W.T.L., Chuang, W.C., and Gober, P., 2012. Vulnerability to extreme heat in metropolitan phoenix: Spatial, temporal, and demographic dimensions. *The Professional Geographer*, 64 (2), 286–302.
- Conlon, K., et al., 2020. Mapping human vulnerability to extreme heat: A critical assessment of heat vulnerability indices created using principal components analysis. *Environmental Health Perspectives*, 128 (9), 97001.
- Cutter, S.L., Boruff, B.J., and Shirley, W.L., 2003. Social vulnerability to environmental hazards. *Social Science Quarterly*, 84, 242–261.
- Dunkel, A., 2015. Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, 142, 173–186.
- Fohringer, J., et al., 2015. Social media as an information source for rapid flood inundation mapping. *Natural Hazards and Earth System Sciences*, 15 (12), 2725–2738., 2015.
- Gao, Y., et al., 2018. Mapping spatiotemporal patterns of events using social media: A case study of influenza trends. *International Journal of Geographical Information Science*, 32 (3), 425–449.
- Goodchild, M.F., 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69 (4), 211–221.
- Goodman, T., Iserson, K.V., and Strich, H., 2001. Wilderness mortalities: A 13-year experience. *Annals of Emergency Medicine*, 37 (3), 279–283.
- Han, S.Y., et al., 2023. A cyberGIS approach to exploring neighborhood-level social vulnerability for disaster risk management. *Transactions in GIS*, 27 (7), 1942–1958.
- Harlan, S.L., et al., 2013. Neighborhood effects on heat deaths: Social and environmental predictors of vulnerability in Maricopa County, Arizona. *Environ. Environmental Health Perspectives*, 121 (2), 197–204.
- Ilieva, R.T., and McPhearson, T., 2018. Social-media data for urban sustainability. *Nature Sustainability*, 1 (10), 553–565.
- Johnson, D.P., et al., 2012. Developing an applied extreme heat vulnerability index utilizing socio-economic and environmental data. *Applied Geography*, 35 (1–2), 23–31.
- Karmegam, D., and Mappillairaju, B., 2020. Spatio-temporal distribution of negative emotions on Twitter during floods in Chennai, India, in 2015: A post hoc analysis. *International Journal of Health Geographics*, 19 (1), 19.
- Kovács, Z., et al., 2021. Combining social media and mobile positioning data in the analysis of tourist flows: A case study from Szeged Hungary. *Sustainability (Sustainability)*, 13 (5), 2926.
- Kumar, V., and Subba, B., 2020. A TfIdfVectorizer and SVM based sentiment analysis framework for text data corpus. In *2020 national conference on communications (NCC)*, 1–6. IEEE.

- Lee, R., Wakamiya, S., and Sumiya, K., 2013. Urban area characterization based on crowd behavioral lifelogs over Twitter. *Personal and Ubiquitous Computing*, 17 (4), 605–620.
- Leetaru, K., et al., 2013. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18 (5). <https://doi.org/10.5210/fm.v18i5.4366>
- Lehnert, E.A., et al., 2020. Spatial exploration of the CDC's Social Vulnerability Index and heat-related health outcomes in Georgia. *International Journal of Disaster Risk Reduction*, 46, 101517.
- Li, H., et al., 2018. Interaction between urban heat island and urban pollution island during summer in Berlin. *The Science of the Total Environment*, 636, 818–828.
- Li, Z., et al., 2018. A novel approach to leveraging social media for rapid flood mapping: A case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science*, 45 (2), 97–110.
- Li, M., et al., 2021. Revitalizing historic districts: Identifying built environment predictors for street vibrancy based on urban sensor data. *Cities*, 117, 103305.
- Li, Y., et al., 2023. Urban flood susceptibility mapping based on social media data in Chengdu city, China. *Sustainable Cities and Society*, 88, 104307.
- Liu, Y.Y., et al., 2018. A CyberGIS integration and computation framework for high-resolution continental-scale flood inundation mapping. *JAWRA Journal of the American Water Resources Association*, 54 (4), 770–784.
- Lugo-Amador, N.M., Rothenhaus, T., and Moyer, P., 2004. Heat-related illness. *Emergency Medicine Clinics of North America*, 22 (2), 315–327.
- Luo, F., et al., 2016. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, 70, 11–25.
- Lyu, F., Yin, D., et al., 2019. Reproducible hydrological modeling with CyberGIS-Jupyter: A case study on SUMMA. In *Proceedings of the practice and experience in advanced research computing on rise of the machines (learning)*, 1–6.
- Lyu, F., et al., 2021. A vector-based method for drainage network analysis based on LiDAR data. *Computers & Geosciences*, 156, 104892.
- Lyu, F., et al., 2022. An integrated CyberGIS and machine learning framework for fine-scale prediction of urban heat island using satellite remote sensing and urban sensor network data. *Urban Informatics*, 1 (1), 6.
- Maier, G., et al., 2014. Assessing the performance of a vulnerability index during oppressive heat across Georgia, United States. *Weather, Climate, and Society*, 6 (2), 253–263.
- Martí, P., Serrano-Estrada, L., and Nolasco-Cirugeda, A., 2019. Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161–174.
- Martin, M.E., and Schuurman, N., 2017. Area-based topic modeling and visualization of social media for qualitative GIS. *Annals of the American Association of Geographers*, 107 (5), 1028–1039.
- McKittrick, M., Schuurman, N., and Crooks, V., 2022. Collecting, analyzing, and visualizing location-based social media data: Review of methods in GIS-social media analysis. *GeoJournal*, 88 (1), 1035–1057.
- Meehl, G.A., and Tebaldi, C., 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science (New York, N.Y.)*, 305 (5686), 994–997.
- Mikolov, T., et al., 2013. Efficient estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR, 2013*.
- Mikolov, T., et al., 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS, 2013*.
- Mislove, A., et al., 2021. Understanding the demographics of Twitter users. *Proceedings of the International AAAI Conference on Web and Social Media*, 5 (1), 554–557.
- Mooney, C.Z., 1997. *Monte Carlo simulation* (No. 116). Thousand Oaks, CA: Sage.
- Muller, C.L., 2013. Mapping snow depth across the West Midlands using social media-generated data. *Weather*, 68 (3), 82–82.

- Niu, Y., et al., 2021. A systematic review of the development and validation of the heat vulnerability index: Major factors, methods, and spatial units. *Current Climate Change Reports*, 7 (3), 87–97.
- Niu, H., and Silva, E.A., 2023. Understanding temporal and spatial patterns of urban activities across demographic groups through geotagged social media data. *Computers, Environment and Urban Systems*, 100, 101934.
- Omuya, E.O., Okeyo, G., and Kimwele, M., 2023. Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. *Engineering Reports*, 5 (3), e12579.
- Padmanabhan, A., et al., 2013. FluMapper: An interactive CyberGIS environment for massive location-based social media data analysis. In *Proceedings of the conference on extreme science and engineering discovery environment: Gateway to discovery*, 1–2.
- Reid, C.E., et al., 2009. Mapping community determinants of heat vulnerability. *Environmental Health Perspectives*, 117 (11), 1730–1736.
- Rong, X., 2016. *word2vec parameter learning explained*.
- Sharma, A., et al., 2016. Green and cool roofs to mitigate urban heat island effects in the Chicago metropolitan area: Evaluation with a regional climate model. *Environmental Research Letters*, 11 (6), 064004.
- Shindell, D., et al., 2020. The effects of heat exposure on human mortality throughout the United States. *GeoHealth*, 4 (4), e2019GH000234.
- Shirowzhan, S., Tan, W., and Sepasgozar, S.M., 2020. Digital twin and CyberGIS for improving connectivity and measuring the impact of infrastructure construction planning in smart cities. *ISPRS International Journal of Geo-Information*, 9 (4), 240.
- Slamet, C., et al., 2018. Social media-based identifier for natural disaster. *IOP Conference Series: Materials Science and Engineering*, 288, 012039.
- Sloan, L., and Morgan, J., 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS One*, 10 (11), e0142209.
- Statista. 2021a. *Distribution of Twitter users worldwide as of April 2021, by age group*. Available from: <https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>
- Statista. 2021b. *Distribution of Twitter users worldwide as of October 2021, by gender*. Available from: <https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/>
- Sui, D., and Goodchild, M., 2011. The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science*, 25 (11), 1737–1748.
- Thorsson, S., et al., 2014. Mean radiant temperature—A predictor of heat related mortality. *Urban Climate*, 10 (2014), 332–345.
- Twitter. 2020. *Twitter API v2 overview to-date: Limit for streaming*. Available from: <https://twitter-community.com/t/limit-for-streaming/138324>
- Twitter. 2020. *Tweet geospatial metadata*. Available from: <https://developer.twitter.com/en/docs/tutorials/tweet-geometadata>
- Twitter. 2021. *Q3 2021 Letter to Shareholders*. Available from: https://s22.q4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3'21-Shareholder-letter.pdf
- Twitter. 2022. *The Twitter Rules*. Available from: <https://help.twitter.com/en/rules-and-policies/twitter-rules>
- Twitter. n.d. *Data dictionary: Standard v1.1*. Available from: <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/geo>
- Van Canneyt, S., et al., 2012. Detecting places of interest using social media. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 447–451.
- Varol, O., et al., 2017. Online human-bot interactions: Detection, estimation, and characterization. *Proceeding International AAAI Conf. on Web and Social Media (ICWSM)*.
- Vishwanath, T., et al., 2023. Social media data extraction for disaster management aid using deep learning techniques. *Remote Sensing Applications: Society and Environment*, 30, 100961.
- Wang, J., et al., 2021. Mapping the exposure and sensitivity to heat wave events in China's megacities. *Science of the Total Environment*, 755 (Pt 1), 142734.

- Wang, S., 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, 100 (3), 535–557.
- Wang, S., 2016. CyberGIS and spatial data science. *GeoJournal*, 81 (6), 965–968.
- Wang, S., et al., 2021. Integrating CyberGIS and urban sensing for reproducible streaming analytics. In: Shi W., Goodchild M.F., Batty M., Kwan MP., Zhang A. eds. *Urban informatics*. Singapore: The Urban Book Series. Springer.
- Xia, C., Schwartz, R., et al., 2014. CityBeat: Real-time social media visualization of hyper-local city data. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. New York, NY, USA: Association for Computing Machinery, 167–170.
- Xue, B., Fu, C., and Shaobin, Z., 2014. A study on sentiment computing and classification of Sina Weibo with Word2vec. *2014 IEEE International Congress on Big Data*, 358–363.
- Yin, D., et al., 2019. CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency and Computation: Practice and Experience*, 31 (11), e5040.
- Zander, K.K., et al., 2023. Responses to heat waves: What can Twitter data tell us? *Natural Hazards*, 116 (3), 3547–3564.
- Zhai, W., et al., 2018. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, 74, 1–12.
- Zhang, K., et al., 2023. Increased heat risk in wet climate induced by urban humid heat. *Nature*, 617 (7962), 738–742.
- Zhao, L., et al., 2014. Strong contributions of local background climate to urban heat islands. *Nature*, 511 (7508), 216–219.
- Zivanovic, S., Martinez, J., and Verplanke, J., 2020. Capturing and mapping quality of life using Twitter data. *GeoJournal*, 85 (1), 237–255.