Advanced Engineering Informatics 50 (2021) 101371

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Crowdsourcing the perceived urban built environment via social media: The case of underutilized land

Yan Wang [a,*], Shangde Gao [b], Nan Li [c], Siyu Yu [d]

[a] Department of Urban and Regional Planning and Florida Institute for Built Environment Resilience, University of Florida, P.O. Box 115706, Gainesville, FL 32611, USA
[b] Department of Urban and Regional Planning, College of Design, Construction and Planning, University of Florida, 1480 Inner Road, Gainesville, FL 32601, USA
[c] Dept. of Construction Management, Tsinghua Univ., Beijing 100084, China
[d] Department of Landscape Architecture & Urban Planning, Texas A&M University, Scoates Hall 125, College Station, TX, USA

ARTICLE INFO

Keywords:
Built environment
Crowdsourcing
Social media
Urban analytics
Underutilized land

ABSTRACT

Crowdsourcing the public's perceptions of the built environment in real time enables more responsive and agile infrastructure and land use planning. Social media has emerged to be an effective platform for citizens, engineers, and planners to communicate opinions and feelings transparently. However, a comprehensive terminological resource of the perceived built environment (BE) for consistent data collection and a specified analytical framework are still lacking, particularly for different underutilized land issues. To fill this knowledge gap, we demonstrate a BE-specific term construction and expansion method specifically for collecting Twitter data and propose a Geo-Topic-Sentiment analytical framework for retrieving and analyzing relevant tweets. We conduct a demonstrative study on un(der)utilized land-related BE terms across ten metropolitan statistical areas in the U.S. Findings reveal spatial variations in contents and sentiments about underutilized land environments, and more localized efforts may be required to address specific land use issues across different urban contexts. The research demonstrates Twitter as a useful platform in crowdsourcing perceived BE and sentiments at fine temporal and spatial scales in a timely manner. It contributes to engineering informatics by investigating the role of social media in environmental planning and proposing integrated domain-specific data analytic approaches for engineering practices.

1. Introduction

The built environment (BE) is at the transformative stage towards being smart, sustainable, healthy, and resilient [17,60]. Cities, as the most populated areas of the world, are facing complex grand challenges including rapid population, environmental issues, aging infrastructure, disasters, diseases, and underutilized land [1,5]. These constant stressors of cities require more responsive, transparent, inclusive, and creative urban planning, especially in the digital era with the emerging Internet of Things. In light of this, communicative planning has been encouraged by citizens, planners, and decision-makers [50]. The key technique to promote more active communicative planning and open innovation is crowdsourcing [50].

Crowdsourcing is an approach that produces data collectively on particular issues, which can help urban managers provide more effective and targeted services at a lower cost and place citizens' needs at the center of thinking and planning [30,8]. Crowdsourcing can also address the research gap in traditional research and management that some place-specific characteristics of distinct urban areas have not been fully considered due to limited data and approaches. To enable crowdsourcing, many participatory platforms and mobile apps have been developed to support citizen participation in urban planning, such as ArcGIS Urban, MindMixer, OpenIDEO, and SafeClickFix. Some apps were designed to collect environmental parameters while others focus on documenting user activities and understanding behaviour [22]. These crowdsourcing models enable more interactive citizen participation in planning, transparent information sharing, and public dialogs [22,50].

In addition to these participatory platforms, social media starts to play a role in the planning process due to the increasing usage of different social networking platforms [51]. Social media has provided a transparent medium for both planners and the general public. It enables direct communications and sentiment expression toward critical procedures during planning. For example, the Department of Planning in Sydney utilized Twitter to engage the community and planning

* Corresponding author.
E-mail addresses: yanw@ufl.edu (Y. Wang), gao.shangde@ufl.edu (S. Gao), nanli@tsinghua.edu.cn (N. Li), syu@arch.tamu.edu (S. Yu).

https://doi.org/10.1016/j.aei.2021.101371
Received 3 March 2021; Received in revised form 30 June 2021; Accepted 3 August 2021
Available online 9 August 2021

authority at key points in the planning process of urban consolidation. The social media simulated detailed debate and contestation over the plan of redevelopment (between local government, a state planning agency, an independent expert panel, journalists, and a self-organized community group) [64]. Planners were suggested to continue the dialogue in a timelier manner [64]. Among the existing most popular social media, Twitter has been mostly studied by planning researchers because of its open design and wide usage. Some studies concerned people's sentiments of a specific aspect of the BE, such as public transit [49], Olympics Games in London [31], and urban green space [47].

Although Twitter has been used as a crowdsourcing platform for different urban topics, few have proposed a complete analytical framework for maximally mining citizens' perceptions of their BE, with a data infrastructure that collects tweets in real time using a comprehensive term list of a BE topic. It is also unclear if Twitter can be a good source for crowdsourcing citizens' perceptions of their BE and how we can retrieve useful information from Tweets to inform urban planners and engineers. No BE terminology list is readily useful for Twitter data collection. Of equal importance is to develop an analytical framework to extract and analyze the relevant and useful information from the crowdsourced data to inform planning practices.

Specifically, we focus on a BE topic: un(der)utilized lands (ULD) as ULDs are potentially risky environmental conditions encountered by millions of community members on a daily basis [5]. ULD is a common urban land use issue [37], and the alteration of vacant land is critical in determining a city's future [38]. ULD can alternatively be associated with redevelopment value and a host of positive externalities (e.g. increasing safety perception) [68]. ULDs are the result of various factors, such as suburbanization, annexation, population loss, disinvestment, or pollution. The investigation of this topic is especially important for community redevelopment and improvements of local economies. Previous studies have examined factors linked to vacant land and abandoned structures using spatial-statistical analysis [37]. Factors in these studies include population and economic changes [28,46], urban expansion [3], and differences in regional locations [4]. However, to understand perceptions of the causes of change in ULD, most studies have adopted a survey method [4,36]; studies applying social media to understand the up-to-date public perceptions toward underutilized urban land are rare. Therefore, this research strives to answer two research questions.

Research question #1(RQ1): Is Twitter a useful source for crowdsourcing public perceptions of the physical built environment (BE), particularly for un(der)utilized lands in the built environment?

Research question #2 (RQ2): What data mining techniques can we employ to formalize the most useful knowledge to inform planning practice?

To answer RQ1, we explored the process to develop a BE terms list of ULDs for collecting tweets and designed a set of algorithms to clean the collected data. Then, to analyze the cleaned dataset, we propose a novel Geo-Topic-Sentiment analytical framework to investigate different concerns and sentiments of ULD-relevant tweets, as well as the spatial differences across ten metropolitan statistical areas (RQ2) in the United States. We present the analytical findings relevant to the un(der)utilized issues in the U.S. and discuss the implications of crowdsourcing people's perceptions with social media data in both urban analytics research and urban planning and engineering practices.

2. Social media and the perceived urban built environment

Studies using social media data to understand human-environment dynamics at an urban scale are burgeoning, such as sensing urban land use [25], exploring urban mobility patterns [69] improving situation awareness in disaster management [20,27], detecting and tracking small-scale urban events [62,65]. These studies have addressed distinct important aspects of smart and responsive cities but not many studies have investigated the usage of social media data in crowdsourcing citizens' perception of their built environment (specifically underutilized lands) and have deep integration with urban and regional planning. We reviewed articles that have used social media data to crowdsource the public's perceptions of their built environment and to inform broad planning practices under the following categories.

2.1. Urban transportation & infrastructure management

Prior studies on social media users' perceptions mostly concerned urban transportation systems, including public transit, metro rail stations, airport service, and general transportation. These studies intended to understand users' sentiments about public services and therefore suggest more effective transportation planning in cities. For example, Schweitzer [49] examined a large sample of Twitter comments about public transit and found more negative sentiments about public transit than sentiments about other public services. This research suggests that planners should involve in social media more actively to foster positive interactions with the public. Spyratos & Stathakis [52] used crowdsourced plan data from Foursquare to evaluate the services and facilities (e.g. sports facilities and streets) of European cities and summarized three aspects of crowdsourcing, including spatially, thematically, and in a timely fashion. Shin [51] examined 830 online reviews of transit users for 54 Los Angeles Metro Rail stations from Yelp. Based on the reviews, the research measured the perceived quality of metro stations and identified the most mentioned factors that influenced user ratings, such as the location of stations, cleanliness, and aesthetic appeal. This study also suggests the importance of building best practices for analyzing social media content. Martin-Domingo et al. [34] analyzed customers' opinions of airport service quality with sentiment analysis of 4392 tweets (obtained from the official Twitter account of London Heathrow airport). They found that waiting, ground transportation, and passport control received more opinions than other service attributes, while the security, check-in, and facilities gain the most satisfactory comments from airport customers. Vasquez-Henriquez et al. [58] analyzed 300 thousand tweets about transportation in Santiago, Chile considering the gender difference. They classified users based on their modes of transportation and estimated the associations between mode, gender, and the categories of a psycho-linguistic lexicon. They identified the differences in the content and sentiment across different traffic modes and genders.

2.2. Landscape evaluation

Researchers from the fields of landscape architecture and urban planning have also used social media to assess citizens' perceptions of urban events or landscapes. Kovacs-Gyori et al. [31] analyzed a set of tweets related to the Olympic Games by examining the sentiments associated with places and differences between residents and visitors. This integration of the new human sensor data can increase the validity and acceptance of governmental decision-making because the data complement traditional planning approaches and reflect citizens' opinions and needs [66]. Roberts et al. [47] conducted a sentiment analysis on 10,000 tweets that are related to urban green space in Birmingham, UK. They found that Tweets are viable data for understanding the human-environment interactions, which can provide valuable information for urban planners and park managers. Researchers from the field of landscape planning have shown particular interest in using crowdsourcing data to analyze and visualize the perceived environment, especially regarding natural resources management. Dunkel [19] conducted a set of landscape perception analyses on crowdsourced photo geodata from Flickr and demonstrated the value of crowdsourcing in integrating public value in the landscape planning process. Brown et al. [6,7] assessed and demonstrated the usefulness of tweets in identifying public perceptions of intrinsic landscape values for landscape design and planning. They mapped the distribution of tweets in the U.S. National Capital Region and discussed the environmental issues and landscape issues as well as the Twitter discussions on the National Climate

AssessmentUnited States Global Change Research Program [57]. Ma et al. [67] analyzed Weibo posts related to urban waterfront spaces in the City of Wuhan in China. By assessing sentiment levels of posts generated nearby the waterfront spaces, they studied the emotion and attitudes of residents surrounding the urban waterfront spaces as well as their spatial distribution. This research proposed that the improvement measures of urban waterfront spaces should be conducted to the locations where Weibo post sentiment was much negative, and also consider the demographic distribution of surrounding residents.

The existing array of studies using social media to understand the perceived BE focuses more on a pre-selected topic (defined by one term). Their analyses either focus on spatial data (e.g. geolocations) or textual data (i.e. contents), but rarely integrate both for crowdsourcing across urban areas. To fill the research gaps, we propose an integrated pipeline of building BE terms for data curation and a data-driven analytical framework to crowdsource individuals' perception and attitudes towards their built environment in real time, which addresses the need of understanding people and their built environment collectively and rapidly.

3. Data and methods

3.1. Generating BETerms for Un(der)utilized lands and building data collection pipeline

This research explores the process to develop a BE-specific term (BETerms) extraction pipeline for the public's perceived built environment. The BETerms is a *comprehensive, specific, and accurate* lexical resource that is suitable for retrieving data from an open-source social media platform (i.e. Twitter). BETerms can be used directly as the keywords query to crowdsource citizens' opinions, sentiments, and comments on their perceived environment in real time and consistently. As BETerms can cover various topics (such as land use, and infrastructure) across disciplines, we specifically focused on the *unutilized and underutilized lands (ULD) BETerms* because vacant and blighted land has been a widespread urban phenomenon in the U.S. and have both negative and positive impacts for cities[37].

Table 1 demonstrates the development of *BETerms* relevant to ULD and the final term list used for collecting tweets. To generate the BETerms, we used two seed terms (Table 1) under the ULD topic. "Vacant" was frequently used when the public discusses issues about unutilized land, and the "blighted area" was used for the un*der*utilized land that caused nearby buildings and/or areas to decline in attractiveness and/or utility. These terms can potentially filter the tweets, which were relevant to the ULD issues. For each seed term, we added all its synonyms and alternative forms based on authoritative reports, academic literature, and disciplinary glossaries, such as *A Planners Dictionary* [14] and *Glossary of Land Use and Planning Terms* [18]. Then we built a Query that includes a comprehensive list of terms used by Twitter users to describe the relevant issues following steps in the below two paragraphs.

Table 1
BETerms under *unutilized and underutilized lands (ULD)* Topics.

Seed terms	Synonymous and alternative forms not in the glossary	Twitter API keywords filters (Directly used as filters in the Twitter API for data collection; underlined terms are newly added)
Vacant, blighted area	Abandoned area, vacant land, blighted land, abandoned land, unused land.	● 'vacant land', 'vacant area', 'vacant parcel', 'urban vacancy', 'urban vacant lots'; ● 'blighted land', 'blighted area'; ● 'abandoned land', 'abandoned area'; ● 'brownfield'; ● 'greyfield'; ● 'polluted land', 'contaminated land'; ● 'unused land'; ● 'undeveloped land'; ● 'empty land'; ● 'unsustainable land';

First, we collected a pilot dataset of tweets using an open-designed, standard Twitter Streaming API [71], which allows at most 400 keywords as the filters[61]. This streaming API can retrieve real-time tweets that contain any filter keyword. Each tweet that has been collected contains user *profile information* (if public), *geographical information, tweet's content* (i.e. *text),* and *time*. Twitter users can share short 140-character messages and the texts may include words, URLs, mentions, emotions, and abbreviations. This dataset was used for evaluating and refining keywords query (i.e. seed terms and alternative forms). If the term in the initial query did not retrieved tweets from the collected tweets, it would be removed from the query; while newly-identified co-occurred high-frequency (top one percent) unigrams (i.e. one-word term) and bi-grams (i.e. two-word term) are added to the query. We assessed and manually refined the query by repeating the process five times to ensure that the terms in the query did not expand since the fourth time. All the newly added terms were underlined in Table 1. This process ensures the data collection query includes as many relevant/alternative terms as it can, but tweets that do not use these terms but contain implicit meanings may be excluded from the final datasets, which is inevitable with current text mining techniques. Finally, we generate the ULD BETerms for consistent Twitter data collection and crowdsourcing opinions through the Twitter Streaming API.

3.2. Geo-topic-sentiment analytical framework

Geo-Topic-Sentiment analytical framework (Fig. 1) focuses on both textual and spatial attributes of tweets. We present the tweets preprocessing, geocoding, and spatial variation analysis of sentiments and contents in the following paragraphs.

Bots removal and tweets preprocessing. Twitter contains an increasing portion of bot accounts [40]. Bots-generated tweets affect the data quality, so we employ a decision-tree based method to identify bots' tweets based on the characteristics of the Twitter accounts, e.g. number of followers/followings, verified accounts or not, number of statuses, number of users' favorite tweets, number of public lists that the user is a member of, and the time when tweets were generated [24]. In our keywords-filtered dataset, the bots' tweets occupy around 39% of all the collected tweets. We also preprocess tweets for further content analysis by removing meaningless and irrelevant components in texts. Specifically, we remove the 'RT' (an abbreviation of "re-tweet") and username mentioning (@username). Then we tokenize words in each tweet so the tweet is stored as a word list. After the tokenization, we remove all numerical numbers, URLs, stopwords (words that have a grammatical function but no meaning, such as "the", "if", "but", and "and"), words that have too long/short length (for words in lowercase the length should more than two, and for words in uppercase the length should not exceed five).

Sentiment analysis is an analytical approach that mines and classifies people's opinions and emotions from textual data into either a numerical scale (e.g. −1 to + 1) or into categories (e.g., positive, negative, or neutral) [61]. Social media data such as tweets contain rich opinions and sentiments of citizens towards various topics [41]. We choose VADER (an abbreviation of Valence Aware Dictionary and Sentiment Reasoner), a well-developed sentiment analysis method, to compute tweets' sentiment scores. VADER is a lexicon and rule-based sentiment analysis method that is specifically attuned to sentiments expressed in social media [29]. VADER can generate continuous numeric sentiment scores of tweets by summing the sentiment value of each word and then normalizing the score to the value between −1.0 to 1.0 (from negative, neutral to positive) [29]. We choose VADER because it has been widely used in previous studies on short texts, e.g. comments on CNN online

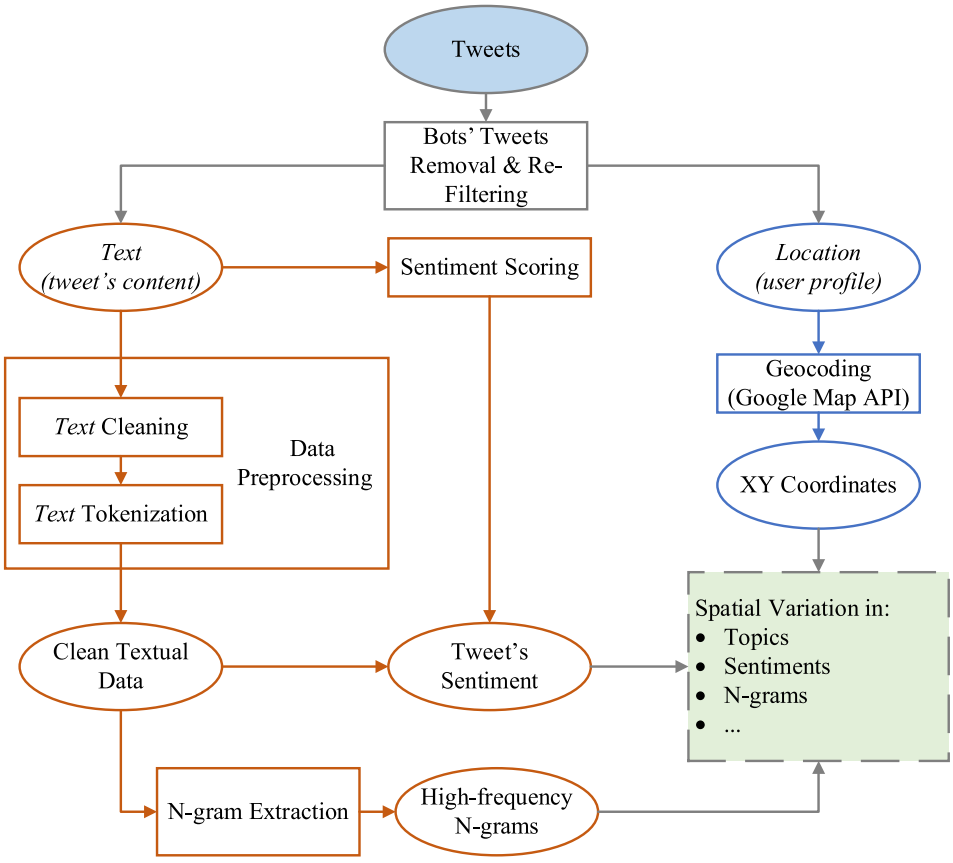

Fig. 1. Geo-Topic-Sentiment Framework for Analyzing BETerms Tweet.

news [9] and tweets [15,23]and more importantly, VADER performs as well as individual human raters at matching ground truth [29].

Geocoding with Google Map API. Location information of Twitter data is stored in different attributes, i.e. *locations* in users' profiles, the *place* tags (e.g. Miami, Florida) that users attach on tweets, or geotagging with specific coordinate information (i.e. latitude and longitude). As we use the keyword filter, the location information is detected in users' profiles and place tags. In our dataset after bot removal, the percentage of tweets containing place tags is 1.22%, which is much lower than the percentage of tweets including the *locations* in users' profiles (~65%). The *locations* of Twitter users are described in diverse ways, e.g., country names, city names, place names, or road names. We employ a set of algorithms developed by Lynn [33] to geocode these *locations* in R Studio, which is also the core algorithm of a Google Map API [26]. The output of the geocoding includes the latitude and longitude of the tweets. The geocoding outputs also indicate differing levels of accuracy (e.g. country, locality, neighborhood, and colloquial area), which is important for deciding the spatial extent when conducting spatial variation analysis across urban areas.

Spatial variation analysis intends to investigate differences in tweets' contents related to BETerms across urban areas. The analysis concerns i) topic variation and ii) sentiment variation. For the topic variation, we regard BETerms as topics, and tweets are divided into distinct specific topics based on if that tweet contains a keyword (regardless of the order of words) that specifies the topic. Then the proportion of tweets under each topic in each area is calculated to identify the most concerned topics. Additionally, we investigate the content of the tweet (i.e. *text* attribute) and extract all the high-frequency N-grams (including unigrams and bigrams [54]) to explore associated terms with topics in much detail. For the sentiment variation across BETerms in distinct areas, we compare the sentiment statistics (such as the distribution and range of sentiment level) across areas to evaluate the aggregated differences in Twitter users' opinions/feelings towards specific BETerms. We also conduct a joint sentiment and N-grams analysis across areas to identify local sentiment tendencies on certain terms and words, for example, terms characterized by the most negative opinions/feelings among local Twitter users.

4. Demonstrative cases, analyses and results

4.1. Beterms for Un(der)utilized lands and Twitter data description

Using the final ULD BETerms, we set up a Twitter streaming API and collected tweets continuously for nearly *two months* in the last quarter of 2019 and retrieved 208,586 tweets in total. We focused on analyzing English tweets because 96.66% of our collected tweets were in English. To clean the dataset, we removed 71,369 tweets that were generated by bot accounts, and the remaining 130,257 tweets were utilized in the next process. We then re-filtered the tweets with the list of terms; if the text part of a tweet did not contain any filter keyword, we removed that tweet. This step is necessary because the Twitter API applies the filter keywords not only in the attribute of *text*, but also in *username*, and *place*. In total, we got 19,105 tweets. Lastly, we geocoded 12,453 tweets that contain useful *location* information. After refining the geocoded Tweets using the boundary of the U.S., we obtained 5,429 ULD-relevant tweets that contain useful U.S. location information for further analysis. We analyzed the two-month dataset for demonstration purposes and intended to show the usefulness of short-term tweets and the approach to mine people's perceptions towards ULD across areas. The distribution of distinct ULD BETerms in the dataset is demonstrated in the pie chart (Fig. 2).

Among all the terms, "brownfield" and "abandoned land/area" occupied half of the dataset. "Vacant land" and "polluted/contaminated land" were also frequently discussed by Twitter users in the U.S. over the

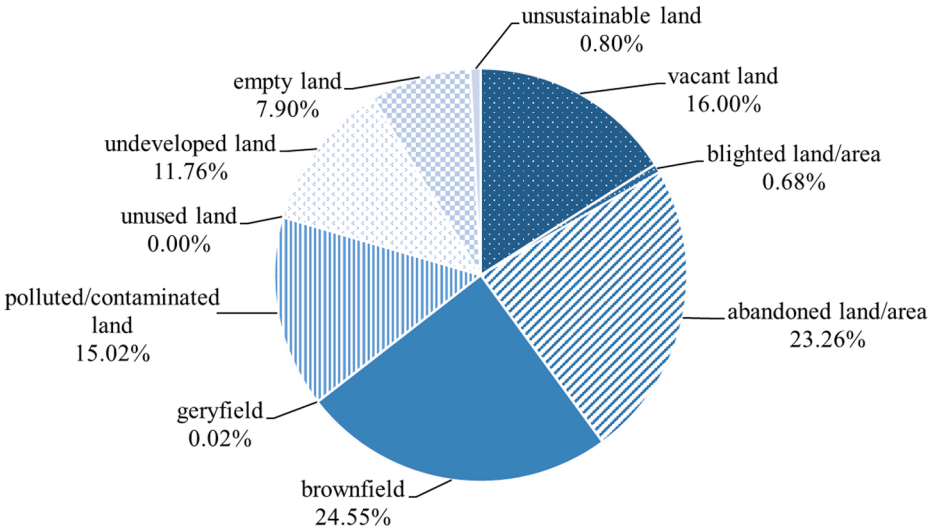

Fig. 2. Percentage of tweets for distinct BETerms (similar terms in Table 1 have been merged).

studied period. The rest of the ULD BETerms (e.g. "unsustainable land", "blighted land/area", "greyfield" and "unused land") have rarely been used in the tweets' dataset in our studied area. Generally, for our study period, Twitter users in the U.S. were concerned more about the (potentially) pollution-caused underutilized land with high percentages of tweets related to contaminated land, polluted land, and brownfield.

4.2. Description of selected urban metropolitan statistical areas

Spatial extent/area selection. To investigate the nuanced differences in the collected tweets across urban areas, we selected ten metropolitan statistical areas (i.e. metros) from the Top 50 metros with the largest population [55]for comparative studies. The ten areas are distributed across five national regions including Northeast, Southwest, West, Southeast, and Midwest [35]. These regions are characterized by different or similar BE conditions. The ten metros also have a large base population and Twitter users. We presented the spatial distribution of the ten metros and their data volumes in Fig. 3. The tweets' data volume varies across metros. For example, New York Metro had the largest dataset with 350 ULD tweets while Miami Metro has the lowest data volume with only 60 tweets. One potential reason underlying the different data volumes is the various local population and the popularity of Twitter.

To compare the content differences across space, we calculated the proportion of tweets related to different terms for each metro. The distribution of the terms for each area is shown in Fig. 4. We found that "vacant/blighted land" topics were more widely discussed in the Metro of Miami, Detroit, Seattle, Chicago, and San Francisco; while Twitter users in Houston and Dallas Metros were concerned more about "polluted/contaminated land". "Abandoned land/areas" and "empty land" were also shared concerns across these areas. With relatively lower percentages, ULD topics, such as "greyfield", "unused land", and "unsustainable land" received different levels of attention across areas.

4.3. Geo-topic-sentiment analysis across metros

The average sentiment levels of ULD tweets in the U.S. Metros during the two months are demonstrated in Fig. 5 and there is no significant spatial pattern identified in the nationwide data. We used a boxplot to compare sentiments on ULD terms across the ten metros (Fig. 6). To

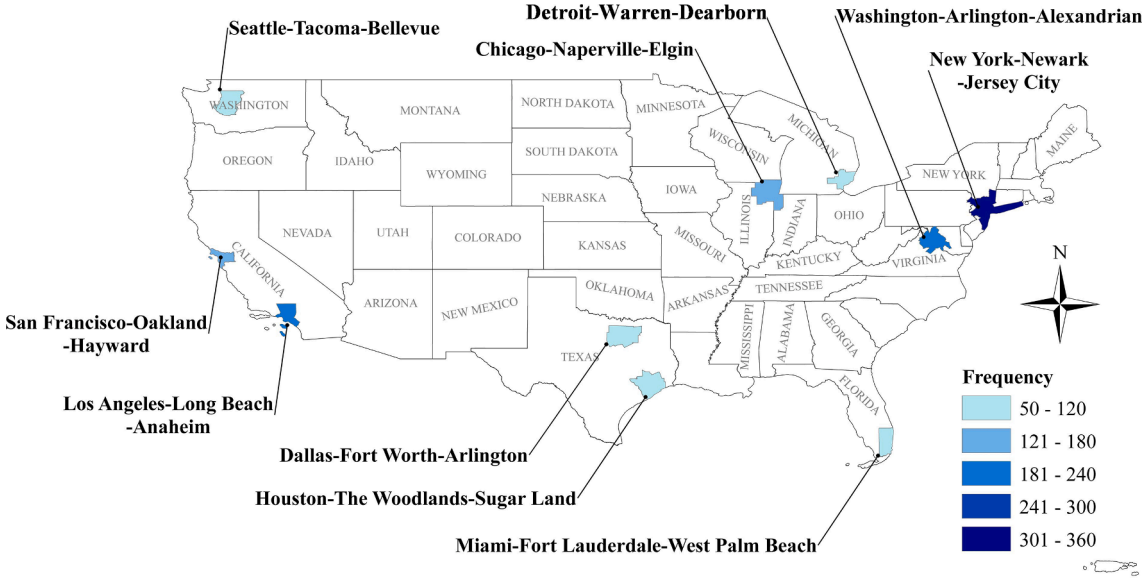

Fig. 3. Spatial distribution and data volume of the ten metros.

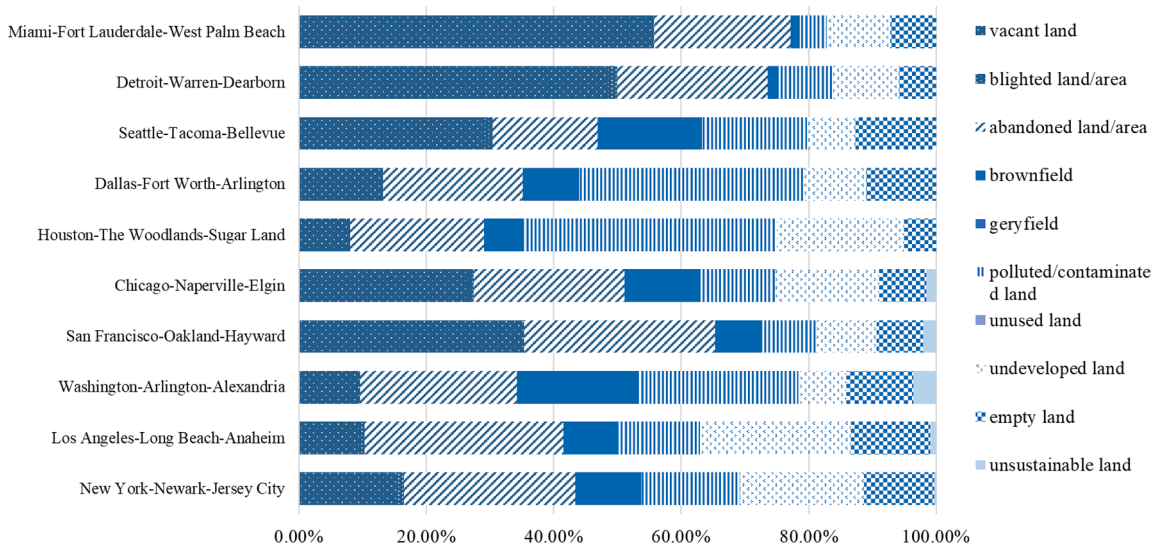

Fig. 4. The proportion of tweets related to different terms in the ten metros.

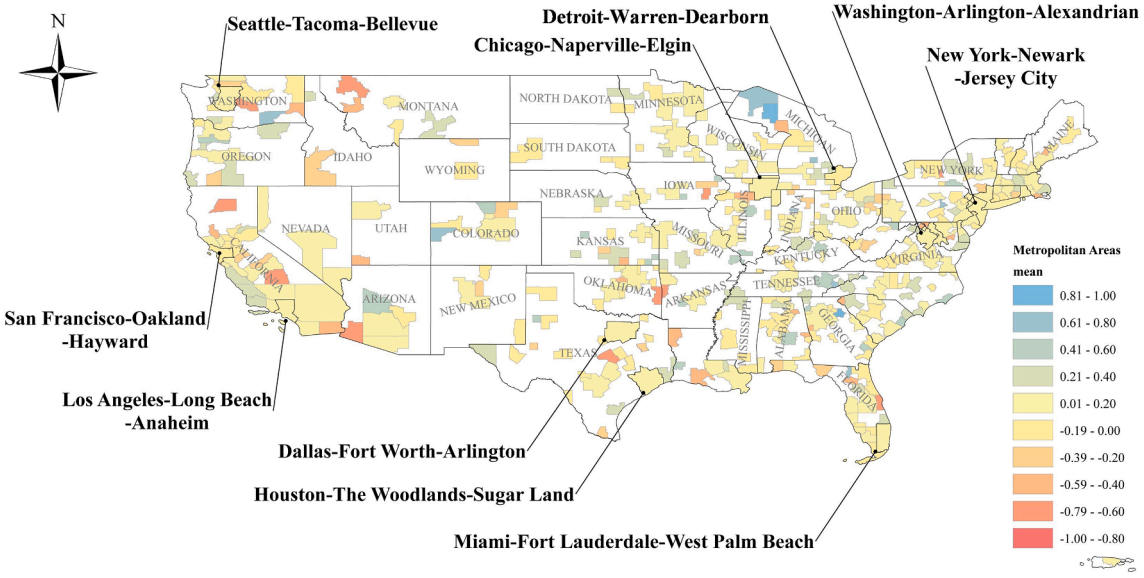

Fig. 5. Mapping sentiment across U.S. metros.

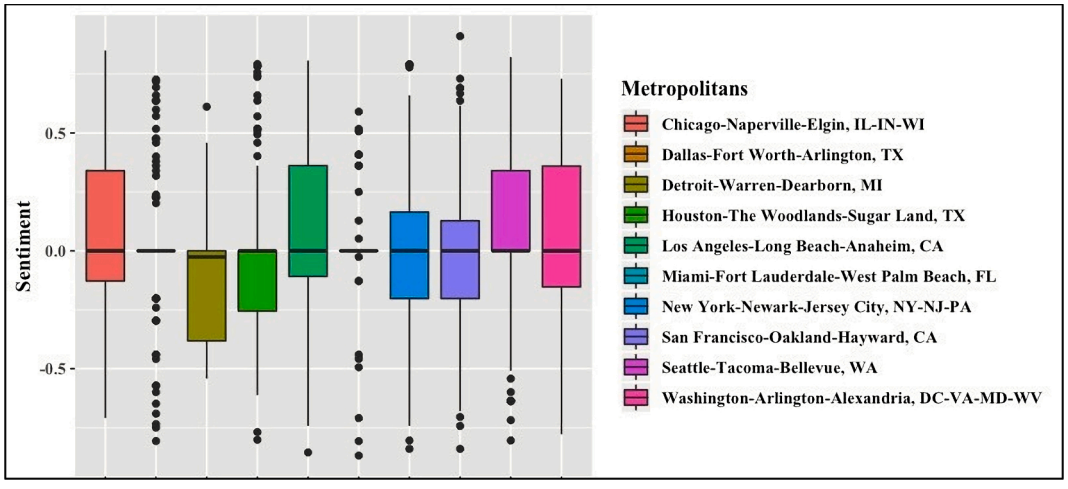

Fig. 6. Sentiments of ULD tweets in the selected ten metros.

validate the accuracy of the VADER sentiment analysis model, we utilized the *simple random sample* method [42] to sample 550 tweets (around 1% of the whole dataset) and annotated them manually, then compared results with another Twitter-specific sentiment analysis model AFINN [39]. The accuracy level of VADER was about 96.4% and outperformed AFINN (94.2%). Miami and Dallas Metros do not have boxes because the data volume in the two metros was relatively small, and their tweets' sentiments spread randomly over the range of sentiment scores, instead of gathering within a certain range. On the contrary, users from Chicago, Los Angeles, Seattle, and Washington Metro posted tweets with a positive tendency; comparatively, the sentiment scores of tweets in Detroit and Houston Metro are mainly distributed over the negative range, indicating that ULD tweets tend to be unsatisfactory. For New York and Los Angeles Metros, sentiment scores of tweets are equally distributed in both positive and negative scales, and the sentiment is gathered around the neutral level.

At the same time, we investigated the usage of words related to the ULD terms in different metros. We extracted all meaningful high-frequency unigrams and bigrams (N-grams thereafter) from the tweets of each metro. To understand the aggregated sentiments of different N-grams for each area's textual dataset, we calculated the average sentiment level associated with each high-frequency N-grams (see Fig. 7). The average sentiment of the N-grams represents the average sentiment of the tweets that contain the unigrams and bigrams in that area. Corresponding to a different data volume of relevant tweets in distinct areas, we were able to identify the 30 top N-grams for New York and Los-Angeles area, 20 top N-grams for Washington, San Francisco, and Chicago Metros areas, and 10 for the rest areas with the relatively smaller corpus size.

On average, Twitter users from different metros have discussions on "*empty (land)*", "*abandoned (land/area)*", "*undeveloped land*", "*vacant land/buildings*" as well as "*contaminated/polluted (land)*" with negative sentiment. For example, the tweet "*Vacant and derelict land can cause long term harm to communities.*" has complained about the negative impact of ULD in the user's local neighborhoods. "*Homeless community/residence*" appears to be a common vacant land-related concern in the most populous metros (e.g. New York, Los Angeles, and San Francisco Metro). It is interesting to find that "*brownfield*" and some ULD topics also appear as a positive word in many areas, especially when they are related to redevelopment and new investment discussions. Notably, N-

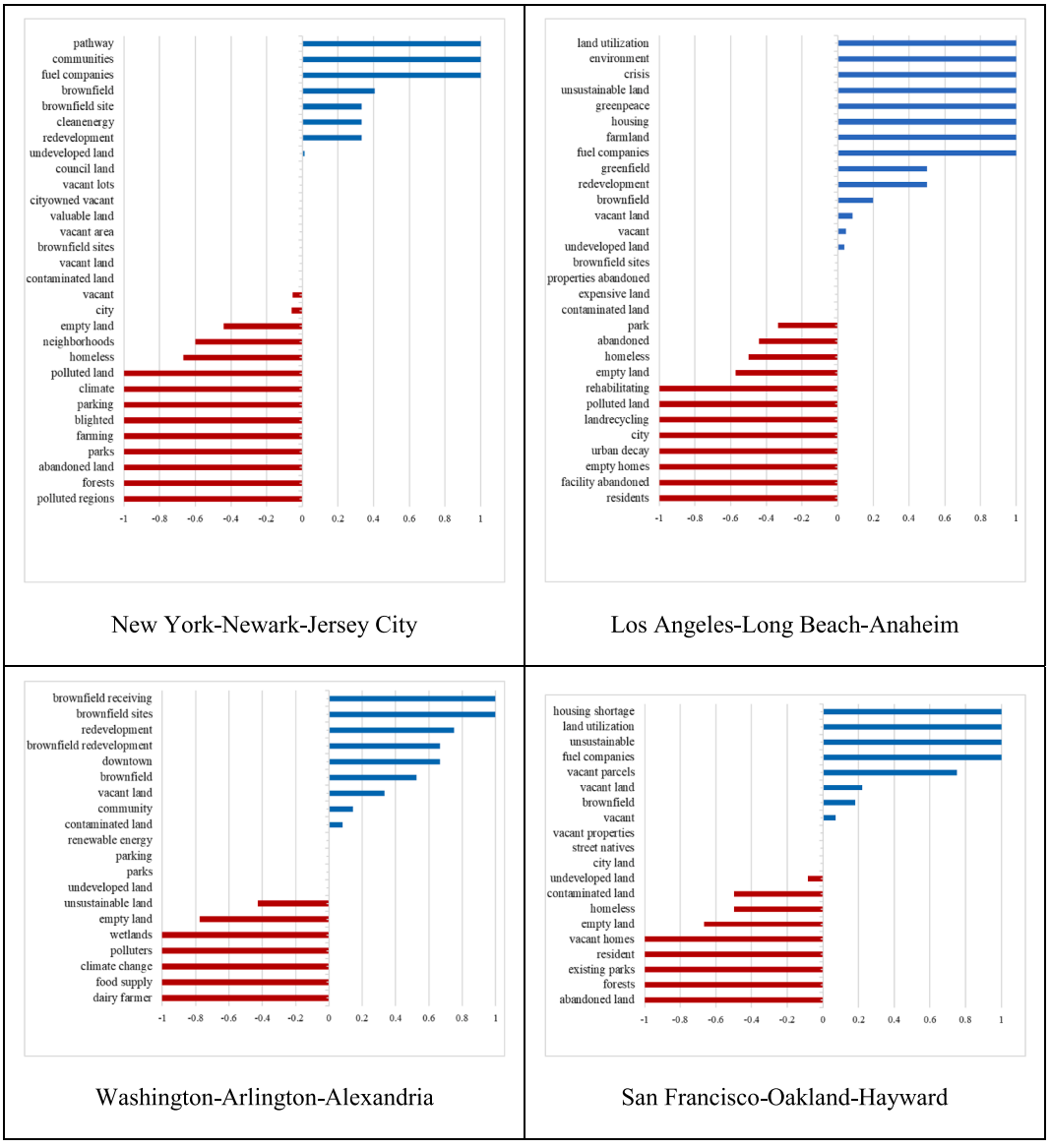

Fig. 7. The average sentiment level of the highest-frequency N-grams. (Red color represents negative sentiment and blue refers to positive sentiments). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

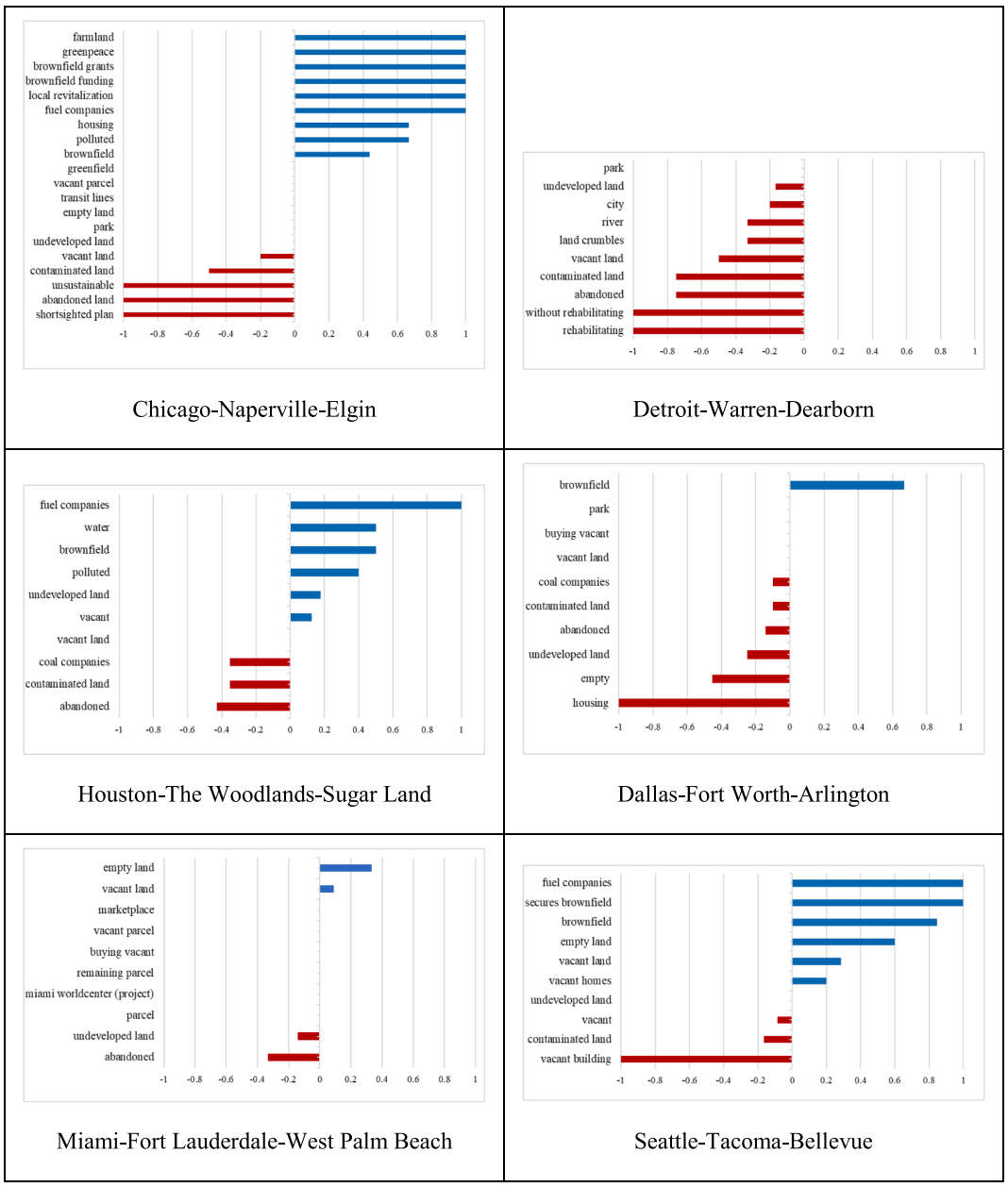

Fig. 7. (*continued*).

grams that are neutral were equally discussed in negative and positive tweets. Two ULD-associated topics have triggered rich discussions across areas in the two months. This is also the reason why some unigrams/bigrams (i.e. coal company and fuel company) had very high frequencies, for example:

"Texas coal companies are leaving behind contaminated land" (from a local Texas news account [12], retweeted frequently nationwide, especially in Houston and Dallas)

"Fossil fuel companies have polluted our air, land, and water for profit – despite knowing the devastating impact it has on…" (Retweeted frequently in Houston, Washington, Dallas, and New York metros)

We find that most metros had both negative and positive tweets for the high-frequency words. For the New York Metro, more positive tweets were related to *"brownfield"*, *"clean energy"* and "*redevelopment"*. Negative tweets also concerned *"homeless"*, "*climate"*, "*farming"*, and "*forests"*. These findings were corresponding to the news about the local land use. For example, New York Times [53] reported that the New York city council provided more affordable housing to mitigate the homeless crisis. In the New York City Strategic Plan, One NYC 2050 identifies opportunities and strategies to achieve carbon neutrality and adapt to climate change to respond with those concerns [11]. In the Washington Metro, most negative words included "*dairy farmer"* (related to a land pollution crisis), "*wetlands"*, *"climate change"*, *"food supply"* (related to the food supply risk caused by unsustainable land use), *"unsustainable land"* and generally, Twitter users had more positive sentiments on "*city"* (related to the redevelopment of vacant land in local cities), "*brownfield/vacant land redevelopment*" (e.g. "*solar panel*", and "*receiving grants*"). Additionally, Washington, D.C. also conducted boards about redeveloping the vacant/abandoned land during the studied periodv [44]. In the Chicago Metro, the shortsighted plan of land use caused very negative opinions while positive N-grams were mostly related to changes in the ULD. A previous study revealed that the City of Chicago owned more than 13,000 vacant lots, and some of them have never been built [2]. For the three selected metropolitans on the West Coast (i.e., Seattle, San Francisco, and Los Angeles area), most of the high-frequency words associated with ULD topics tended to be negative.

ULD concerns also included "*land recycling*", "*urban decay*", "*empty homes*", "*facility abandoned*" (LA); "*parks*" *and* "*forests*" (SF); and "*vacant building*" (Seattle). The potential reason for the frequent uses of these terms was that the 2019 Brownfields National Training Conference in California, emphasized the need for land recycling and redevelopment of vacant and abandoned areas [21]. The Miami Metro area had the lowest data volume, most negative words were "*abandoned* and *undeveloped (land)*" while other high-frequency words had neutral or positive discussions on average. These N-grams include "*(remaining) parcel*", "*Miami worldcenter*", and "*marketplace*". The Houston and Dallas Metros, the two largest metropolitan areas in Texas, shared similar concerns about "*contaminated land*", "*housing*", "*undeveloped land*", "*abandoned land*" and additionally, the "*coal company*", a seeming indictment of BE and land use policy in the state regarding these key issues. This could indicate (or lead, in the future, to) negative perceptions of these metros, and should, therefore, be prioritized among decision-makers. The high-frequency words also matched the news about Texas coal companies, which caused many land pollutions issues. *Plan Houston* [10] also identified these issues and included policies targeting undeveloped land and housing: "The City of Houston plans to adopt more housing policies to support underdeveloped communities and to maintain affordable housing supply" [10].

However, Twitter users from the Detroit Metro area had negative sentiments on average, for all the high-frequency N-grams, such as "*abandoned*", "*contaminated (land)*", "*vacant land*", and "*undeveloped (land)*", as well as associated terms like "*rehabilitating*", "*land crumbles*", "*river*", "*city*", and "*park*". For half of the words, the negative values were relatively strong (lower than −0.5), indicating more complaints and dissatisfaction among the public on ULD issues in the urban area. This result corresponds to findings in other studies that Detroit exemplified the effects of urban vacancy, showing the results of shrinkage [36], and suggests that more policies and design solutions should be built into local plans to address the negative public perceptions towards underutilized lands in the area.

In general, our framework effectively identified the local land use issues of the ten Metros in this study. Across the ten U.S. metropolitan statistical areas, spatial variations in the contents and sentiments about underutilized land reveal differences in the complex issues and experience with which cities in various development stages must content. More localized planning efforts are needed to address specific land use issues across the different urban areas, as reflected on social media.

5. Discussion

Twitter data is a useful medium for crowdsourcing public perceptions of the physical built environment (BE), particularly for un(der) utilized lands at finer temporal and spatial scales, including issues around brownfields, undeveloped land, and vacant land. By creatively dosing data analytics tools across data preprocessing, geocoding, and natural language processing and developing a domain-specific keywords query, the Geo-Topic-Sentiment analytical framework advanced previous studies that relied on surveys to understand public perceptions and opinions with the high cost and low response rate and lack of timeliness [4,36]. Using the collective individuals' opinions, local planners and engineers can translate the knowledge into actions by addressing the identified land use issues in local plans, developing green infrastructure in vacant urban lands, and promote the public's subjective well-being and urban sustainability. Specifically, after identifying social media discourses on land use issues, local planners can verify situations and propose planning strategies to re-develop the underutilized lands. The process for developing BETerms and the analytical framework enables the agile data collection from social media as well as a timely and bottom-up understanding of the public's perceptions of their built environment. Twitter can serve as a supplementary information gathering and discussion tool to collect public ideas and perceptions, accompanying surveys or other community engagement activities over different decision-making stages. The active engagement of planners, researchers, and citizens on social media and neighborhood networking platforms can facilitate more transparent communications of environmental issues and redevelopment plans [70].

As voluntarily reported data, tweets collected for different areas over distinct periods may have differing data bias issues, one of which is the population representation (e.g. younger people and English speakers tweeting more). The representativeness of Twitter users can vary from case to case. This research does not intend to identify empirical findings that can be generalized to the overall population; instead, it focuses on developing the research framework for using Twitter as a supplementary data source for existing crowdsourcing methods in urban planning [30,8]. It is also challenging to infer the demographic characteristics of Twitter users from their online profiles. Thus, this study did not evaluate the demographic characteristics of Twitter users by comparing Twitter users to statistics from the American Community Survey [56]. For future empirical studies that seek representative and generalized findings, we suggest employing a census-derived weighting mechanism if any bias is identified between the ratio of Twitter users and the true population structure (e.g. [13]; Wang et al. 2018). Future research can also extend the work to other popular social media platforms such as Weibo but needs to understand the differences in platform functions and user structures.

Raw tweets also contain noisy data. To ensure the quality of the tweets' content, we focused on removing bots accounts using the BotsOrNot system [16]. We also filtered tweets based on a set of strict standards (e.g. if terms appear in the main text and if tweets having location attributes and are inside areas of interest) and cleaned the data to retrieve useful/relevant information for content and sentiment analysis. Future studies could expand the preprocessing process of social media posts by leveraging machine learning methods for checking the data's credibility as fake news, misinformation, and disinformation coexist with factual information in the complex online environment [45,59]. Such information may affect the identification of credible issues from big data. As we did not detect any dominant rumors in our pilot dataset, this step was not included for ULD data preprocessing.

Regarding the proposed analytical framework for analyzing BETerm-relevant tweets, existing unsupervised text mining algorithms (e.g. [29] have not considered domain-specific features of data, especially in most urban studies. This can affect the analysis accuracy when a word expresses a different meaning used in distinct contexts. For example, although "*brownfield*" occurs in many tweets, it sometimes refers to a technical term meaning "problem and the process of having to consider already existing systems when implementing new software systems" instead of a land-use term [14]. As the deployed unsupervised algorithms are not able to clean the tweets including the "*brownfield*" of the different meanings, we reviewed the datasets of the ten metros and manually removed these irrelevant tweets. Future methodology development research for urban analytics can be improved by enabling more domain-specific analyses.

In addition to the textual analysis, the spatial information retrieved from the *location* attribute of tweets is the other critical component of our proposed analytical framework. We only used the keywords filter in the Twitter API to collect real-time tweets in a stream because additional filters, e.g. geotagging, are not allowed for a free Twitter API, and using two filers can generate a lower percentage of tweets due to the fact that only around 1% of tweets were geotagged [63]. Thus, we employed a Google Map API to conduct the geocoding (i.e. inferring XY locations of the tweets based on the textual location). However, the *location* in the users' profile mostly represents the locality of the Twitter users and may not be the metros where the tweets were generated or about. This means that visitors' perceptions, whose Twitter profiles have different locations, may be left out. For users that do not have profile locations, it is also challenging to retrieve semantically hidden or unstructured spatial information referred to by the text. A spatial information mining technique based on multimodal data may be needed to address the

limitation. However, our strict data preprocessing ensures the reliability of the data used for demonstrative analyses.

Another challenge in crowdsourcing the perceived BE with social media is interpreting outcomes of sentiment analysis as citizens' feelings and opinions. Existing sentiment analysis algorithms are not yet able to detect underlying feelings of tweets. A recent study argues that sentiment analysis of tweets based on affective words (emotional languages) may not be a "perfect mirror of experienced emotion" [32], pp20) because tweets may dampen the affective intensity. Still, some studies have shown that social media posts can deliver individuals' perceptions and opinions towards specific issues [43,48]. In light of this, researchers and planners need to be careful when interpreting the sentiments of users as the actual attitude to specific events or topics, because the emotional intensity may be affected when the users generate tweets after perceiving the related events. However, in this study, sentiment analysis is still useful in terms of the comparisons of N-grams and aggregated contents across areas and terms. Without too much focus on the specific sentiment intensity level, the general trend of negative or positive can still indicate the public's opinions on the land use issues (related to certain BETerms) effectively.

6. Conclusion

Social networking platforms have provided enormous opportunities for crowdsourcing and analyzing citizen's perceptions and feelings of their built environment. To address the research needs of developing comprehensive domain-specific terms for accurate data collection and analysis to inform urban planning and environmental engineering, this research proposed the BE-specific term construction and expansion method and the Geo-Topic-Sentiment analytical framework for retrieving useful spatial and textual knowledge from Twitter data. The demonstrative analysis of ULD terms across ten U.S. metropolitan areas reveals the spatial variations in contents and sentiments of ULD topics. This research finds that Twitter is a useful information source in terms of crowdsourcing built environment perception and formalizing knowledge for decision makings but requires careful and well-designed data cleaning, preprocessing, and examination. Although our study focuses on Twitter data, the research framework is also applicable to other social sensing data when it becomes available. The analytical framework can facilitate various urban planning tasks by quickly "sensing" the public opinion and provide data-driven avenues to inform infrastructure and habitat planning. It would also provide a novel, convenient and informatics-driven means for public engagement and encourage deeper community engagement in the decision-making process. To build responsive, resilient, and smart built environments, crowdsourcing citizens' perceptions require more proactive and domain-specific urban analytics and knowledge formation processes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Science Foundation under Grant No. 1951816 and the University of Florida DCP Research Seed Grant Initiative. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and the University of Florida.

References

[1] American Society of Civil Engineers (ASCE). (2017). America's Infrastructure Report Card 2017. Retrieved February 29, 2018. https://www.infrastructurereportcard.org/.
[2] E.C. Anderson, E.S. Minor, Assessing social and biophysical drivers of spontaneous plant diversity and structure in urban vacant lots, Sci. Total Environ. 653 (2019) 1272–1281, https://doi.org/10.1016/j.scitotenv.2018.11.006.
[3] A. Berger, Drosscape: Wasting Land Urban America, Princeton Architectural Press, New York, 2007.
[4] Bowman, A. O. M., and Pagano, M. A. (2000). Transforming America's cities: Policies and conditions of vacant land. Urban Affairs Review, 35(4), 559-581. https://doi.org/ 10.1177/10780870022184534.
[5] C.C. Branas, E. South, M.C. Kondo, B.C. Hohl, P. Bourgois, D.J. Wiebe, J. M. MacDonald, Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear, Proc. Natl. Acad. Sci. 115 (12) (2018) 2946–2951, https://doi.org/10.1073/pnas.1718503115.
[6] M. Brown, T. Murtha, Y. Wang, L. Wang, ILAS: intrinsic landscape assessment system for landscape design and planning in the national capital region, Journal of Digital Landscape Architecture (2019) 84–94. https://doi.org/10.14627/537663009.
[7] M. Brown, T. Murtha, L. Wang, Y. Wang, Mapping Landscape Values with Social Media, Journal of Digital Landscape Architecture (2020) 542–548.
[8] F. Caprotti, Future cities: moving from technical to human needs. *Palgrave*, Communications 4(1), 35 (2018) 1–4, https://doi.org/10.1057/s41599-018-0089-5.
[9] Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 1217-1230. https://doi.org/10.1145/2998181.2998213. https://doi.org/10.1145/2998181.2998213.
[10] City of Houston. (2015). Plan Houston. Retrieved from: http://www.houstontx.gov/planning/GeneralPlan/generalplan.html.
[11] City of New York. (2019). OneNYC 2050: Building a Strong and Fair City. Retrieved from: https://onenyc.cityofnewyork.us.
[12] Collier, K. (2019). Texas Coal Companies Are Leaving Behind Contaminated Land. The State Is Letting Them. Retrieved from https://www.texastribune.org/2019/10/30/how-texas-lets-coal-companies-leave-behind-contaminated-land/?utm_campaign=trib-socialandutm_content=1572402048andutm_medium=socialandutm_source=twitter.
[13] A. Culotta, Reducing sampling bias in social media data for county health inference, Joint Statistical Meetings Proceedings (2014) 1–12.
[14] M. Davidson, F. Dolnick, A Planner's Dictionary, American Planning Association Planning Advisory Service, Chicago, Michigan, 2004 https://www.planning.org/publications/report/.
[15] T. Davidson, D. Warmsley, M. Macy, I. Weber, Automated hate speech detection and the problem of offensive language, Eleventh international AAAI conference on web and social media (2017) 512–515.
[16] C.A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, Botornot: A system to evaluate social bots, in: Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 273–274, https://doi.org/10.1145/2872518.2889302.
[17] R. DesRoches, J. Taylor, The promise of smart and resilient cities, The Bridge 48 (2) (2018) 13–20.
[18] T. Duarte, M. Stepner, J.J. Roberts, C. Dickson, J. Rosenhall, Understanding the Basics of Land Use and Planning: Glossary of Land Use and Planning Terms. The Institute for Local Government, 2010.
[19] A. Dunkel, Visualizing the perceived environment using crowdsourced photo geodata, Landscape Urban Plann. 142 (2015) 173–186, https://doi.org/10.1016/j.landurbplan.2015.02.022.
[20] N. Eligüzel, C. Çetinkaya, T. Dereli, Comparison of different machine learning techniques on location extraction by utilizing geo-tagged tweets: A case study, Adv. Eng. Inf. 46 (2020) 101151, https://doi.org/10.1016/j.aei.2020.101151.
[21] EPA. (2019). EPA's 2019 National Brownfields Training Conference Wraps Up in Los Angeles California. Retrieved on June 3, 2021. Retrieved from https://www.epa.gov/newsreleases/epas-2019-national-brownfields-training-conference-wraps-los-angeles-california.
[22] T.P. Ertiö, Participatory apps for urban planning—space for improvement, Planning Practice and Research 30 (3) (2015) 303–321, https://doi.org/10.1080/02697459.2015.1052942.
[23] R. Fan, O. Varol, A. Varamesh, A. Barron, I.A. van de Leemput, M. Scheffer, J. Bollen, The minute-scale dynamics of online emotions reveal the effects of affect labeling, Nat. Hum. Behav. 3 (1) (2019) 92–100, https://doi.org/10.1038/s41562-018-0490-5.
[24] E. Ferrara, Disinformation and social bot operations in the run up to the 2017 French presidential election, First Monday 22 (8) (2017) 1–33, https://doi.org/10.5210/fm.v22i8.8005.
[25] V. Frias-Martinez, E. Frias-Martinez, Spectral clustering for sensing urban land use using Twitter activity, Eng. Appl. Artif. Intell. 35 (2014) 237–245, https://doi.org/10.1016/j.engappai.2014.06.019.
[26] Google. (2019). Google Maps Platform Terms of Service. Retrieved from https://cloud.google.com/maps-platform/terms/.
[27] Haiyan Hao, Yan Wang, Leveraging multimodal social media data for rapid disaster damage assessment, Int. J. Disaster Risk Reduct. 51 (2020) 101760, https://doi.org/10.1016/j.ijdrr.2020.101760.

[28] M. Heckert, J. Mennis, The economic impact of greening urban vacant land: a spatial difference-in-differences analysis, Environment and Planning A 44 (12) (2012) 3010–3027, https://doi.org/10.1068/a4595.
[29] C.J. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, Eighth international AAAI conference on weblogs and social media (2014) 216–225.
[30] B. Klein, R. Koenig, G. Schmitt, Managing urban resilience, Informatik-Spektrum 40 (1) (2017) 35–45, https://doi.org/10.1007/s00287-016-1005-2.
[31] A. Kovacs-Gyori, A. Ristea, C. Havas, B. Resch, P. Cabrera-Barona, #London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data, Urban Planning 3 (1) (2018) 75–99. https://doi.org/10.17645/up.v3i1.1287.
[32] M.D. Lieberman, Affect labeling in the age of social media, Nat. Hum. Behav. 3 (1) (2019) 20–21, https://doi.org/10.1038/s41562-018-0487-0.
[33] Lynn, S. (2013). Batch Geocoding with R and Google Maps. Retrieved from https://www.shanelynn.ie/massive-geocoding-with-r-and-google-maps/.
[34] L. Martin-Domingo, J.C. Martín, G. Mandsberg, Social media as a resource for sentiment analysis of Airport Service Quality (ASQ), Journal of Air Transport Management 78 (2019) 106–115, https://doi.org/10.1016/j.jairtraman.2019.01.004.
[35] National Geographic Society. (2012). United States Regions. Retrieved December 30, 2019, from https://www.nationalgeographic.org/maps/united-states-regions/#us-regions-map-1.
[36] G.D. Newman, A.O.M. Bowman, R. Jung Lee, B. Kim, A current inventory of vacant urban land in America, Journal of Urban Design 21 (3) (2016) 302–319, https://doi.org/10.1080/13574809.2016.1167589.
[37] Galen Newman, Yunmi Park, Ann O'.M. Bowman, Ryun Jung Lee, Vacant urban areas: Causes and interconnected factors, Cities 72 (2018) 421–429, https://doi.org/10.1016/j.cities.2017.10.005.
[38] G. Newman, J. Saginor, Four imperatives for preventing demolition by neglect, Journal of Urban Design 19 (5) (2014) 622–637, https://doi.org/10.1080/13574809.2014.943705.
[39] Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.
[40] A. Obadimu, E. Mead, S. Al-Khateeb, N. Agarwal, A comparative analysis of Facebook and Twitter bots, Proceedings of the Southern Association for Information Systems Conference 25 (2019) 1–6.
[41] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, LREC 10 (2010) (2010) 1320–1326.
[42] Pandas. (2021). Pandas. Retrieved on June 16, 2021. Retrieved from https://pandas.pydata.org/.
[43] L. Qiu, H. Lin, J. Ramsay, F. Yang, You are what you tweet: Personality expression and perception on Twitter, J. Res. Pers. 46 (6) (2012) 710–718, https://doi.org/10.1016/j.jrp.2012.08.008.
[44] Rani, R. S. (2019). Washington, D.C., Aims to Improve Underserved Neighborhoods Through Better Land Management. Retrieved on June 3, 2021. Retrieved from https://nextcity.org/daily/entry/washington-dc-aims-to-improve-neighborhoods-through-better-land-management.
[45] Julio C.S. Reis, Andre Correia, Fabricio Murai, Adriano Veloso, Fabricio Benevenuto, Erik Cambria, Supervised learning for fake news detection, IEEE Intell. Syst. 34 (2) (2019) 76–81, https://doi.org/10.1109/MIS.967010.1109/MIS.2019.2899143.
[46] T. Rieniets, Shrinking cities: Causes and effects of urban population losses in the twentieth century, Nature and Culture 4 (3) (2009) 231–254, https://doi.org/10.3167/nc. 2009.040302.
[47] H. Roberts, J. Sadler, L. Chapman, The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation, Urban Studies 56 (4) (2019) 818–835, https://doi.org/10.1177/0042098017748544.
[48] A.B. Rosenkrantz, A. Labib, K. Pysarenko, V. Prabhu, What do patients tweet about their mammography experience? Acad. Radiol. 23 (11) (2016) 1367–1371, https://doi.org/10.1016/j.acra.2016.07.012.
[49] Schweitzer, L. (2014). Planning and social media: a case study of public transit and stigma on Twitter. Journal of the American Planning Association, 80(3), 218-238. https://doi.org/10.10 80/01944363.2014.980439.
[50] E. Seltzer, D. Mahmoudi, Citizen participation, open innovation, and crowdsourcing: Challenges and opportunities for planning, Journal of Planning Literature 28 (1) (2013) 3–18, https://doi.org/10.1177/0885412212469112.
[51] E.J. Shin, What can we learn from online reviews? Examining the reviews of Los Angeles metro rail stations, Journal of Planning Education and Research 0739456X19870261 (2019), https://doi.org/10.1177/0739456X19870261.
[52] S. Spyratos, D. Stathakis, Evaluating the services and facilities of European cities using crowdsourced place data, Environment and Planning B: Urban Analytics and City Science 45 (4) (2018) 733–750, https://doi.org/10.1177/0265813516686070.
[53] Stewart, N., Mays, J.C., & Haag, M. (2019). Facing Homeless Crisis, New York Aims for 1,000 New Apartments a Year. Retrieved on June 1, 2021. Retrieved from https://www.nytimes.com/2019/12/12/nyregion/homeless-housing-nyc.html.
[54] C.M. Tan, Y.F. Wang, C.D. Lee, The use of bigrams to enhance text categorization, Inf. Process. Manage. 38 (4) (2002) 529–546, https://doi.org/10.1016/S0306-4573(01)00045-0.
[55] United States Census Bureau. (2019). Population of the Largest Metropolitan Areas in the U.S. As of 2018 (In 1,000s). Chart. Accessed January 21, 2020. https://www.statista.com/statistics/183600/population-of-metropolitan-areas-in-the-us/.
[56] United States Census Bureau's American Community Survey Office. (2018). American Community Survey 2013-2017 5-year Data Release. Accessed December 6, 2018. https://www.census.gov/newsroom/press-kits/2018/acs-5year.html.
[57] United States Global Change Research Program. (2018). 4th National Climate Assessment. Accessed November 23, 2018. https://nca2018.globalchange.gov/.
[58] P. Vasquez-Henriquez, E. Graells-Garrido, D. Caro, Characterizing transport perception using social media: differences in mode and gender, in: Proceedings of the 10th ACM Conference on Web Science, 2019, pp. 295–299, https://doi.org/10.1145/3292522.3326036.
[59] Yan Wang, Haiyan Hao, Lisa Sundahl Platt, Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter, Comput. Hum. Behav. 114 (2021) 106568, https://doi.org/10.1016/j.chb.2020.106568.
[60] Yan Wang, David Hulse, Jason Von Meding, Madeline Brown, Laura Dedenbach, Conceiving resilience: Lexical shifts and proximal meanings in the human-centered natural and built environment literature from 1990 to 2018, Developments in the Built Environment 1 (2020) 100003, https://doi.org/10.1016/j.dibe.2019.100003.
[61] Y. Wang, J.E. Taylor, Coupling sentiment and human mobility in natural disasters: A Twitter-based study of the 2014 South Napa Earthquake, Nat. Hazards 92 (2) (2018) 907–925, https://doi.org/10.1007/s11069-018-3231-1.
[62] Y. Wang, J.E. Taylor, DUET: Data-driven approach based on latent dirichlet allocation topic modeling, J. Comput. Civil Eng. 33 (3) (2019) 04019023, https://doi.org/10.1061/(ASCE)cp.1943-5487.0000819.
[63] Yan Wang, Qi Wang, John E. Taylor, Wei-Xing Zhou, Aggregated responses of human mobility to severe winter storms: An empirical study, PLoS ONE 12 (12) (2017) e0188734, https://doi.org/10.1371/journal.pone.0188734.
[64] W. Williamson, K. Ruming, Urban consolidation process and discourses in Sydney: Unpacking social media use in a community group's media campaign, Planning Theory and Practice 18 (3) (2017) 428–445, https://doi.org/10.1080/14649357.2017.1328768.
[65] F. Yao, Y. Wang, Tracking urban geo-topics based on dynamic topic model, Comput. Environ. Urban Syst. 79 (2020), 101419, https://doi.org/10.3390/su9122158.
[66] Zeile, P., Resch, B., Exner, J. P., and Sagl, G. (2015). Urban emotions: benefits and risks in using human sensory assessment for the extraction of contextual emotion information in urban planning. Planning Support Systems and Smart Cities. 209-225. Springer, Cham.
[67] Y. Ma, C. Ling, J. Wu, Exploring the spatial distribution characteristics of emotions of Weibo users in Wuhan waterfront based on gender differences using social media texts, ISPRS Int. J. Geo-Inform. 9 (8) (2020) 465.
[68] B.W. Stanley, Leveraging public land development initiatives for private gain: The political economy of vacant land speculation in Phoenix, Arizona, Urban affairs Rev. 52 (4) (2016) 559–590.
[69] Q. Wang, N.E. Phillips, M.L. Small, R.J. Sampson, Urban mobility and neighborhood isolation in America's 50 largest cities, Proc. Natl. Acad. Sci. 115 (30) (2018) 7735–7740.
[70] M.N. Bakht, T.E. El-Diraby, M. Hosseini, Game-based crowdsourcing to support collaborative customization of the definition of sustainability. Adv. Eng. Informat. 38 (2018) 501–513.
[71] 2021. https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/overview. (Accessed 1 July 2021).