Green tweets: evaluating the potential of Twitter in health-related green space research.

March 6, 2017

Submitted by:

Andrew Larkin

larkinan@gwmail.gwu.edu

in Partial Fulfillment of the Requirements for the Master of Public Health Degree, Milken Institute School of Public Health, The George Washington University

**Abstract**

Purpose: Green space is associated with myriad health outcomes, including increased longevity, cognitive function, and social cohesion. These benefits can mitigate unique health challenges young adults face today, including increased levels of stress and obesity. Social media, including Twitter, is used by 85-90% of young adults in the US, and can potentially facilitate public health efforts to research and promote young adult green space attitudes, perceptions, and use.

Methods: A literature review was performed using PubMed, Scopus, Google Scholar, and PsycINFO search engines. A 48-hour sample of Twitter posts (tweets) from Portland, OR containing green space-related keywords w collected and encoded in NVivo for sentiment and thematic analysis. Tweets were also analyzed using automated scripts provided by the Stanford CoreNLP.

Results: Of the 18 papers from the literature review that met inclusion criteria, only one paper utilized social media for green space research. Previous studies found that green space attitudes, use, and perceptions are dependent on local context. Among the 602 collected tweets, many provided time- and location- specific contextual details. Sentiment was significantly more positive among the tweets (7%) with green-space related hashtags compared to the overall sample. Episodic inclement weather during sampling limited the number and generalizability of health promotion/prevention related green space tweets.

Conclusions: Social media can contribute to research knowledge of local green space attitudes, use, perceptions, and context. However, hashtag-based sampling may lead to significant sampling bias, and limitations such as ecological fallacy and seasonal effects must be considered during study design and analysis.

**Background**

Natural components of urban environments, also called green spaces, are associated with multiple positive health outcomes. These outcomes span across multiple dimensions of human health for those living in close proximity. Biological health outcomes include increased birth weight (Ebisu, Holford, & Bell, 2016), decreased blood pressure (Shanahan et al., 2016), decreased risk of type 2 diabetes (Astell-Burt, Feng, & Kolt, 2014), rates of preschool childhood obesity (Lovasi et al., 2013), and mortality rate among senior citizens (Takano, Nakamura, & Watanabe, 2002). Mental health outcomes include increased cognitive function (Bratman, Hamilton, & Daily, 2012), ability to cope with stress (van den Berg, Maas, Verheij, & Groenewegen, 2010), increased positive attitude (Takano et al., 2002), and lower levels of symptoms associated with depression and anxiety (Beyer et al., 2014). Community health outcomes include increased social cohesion (Shanahan et al., 2016), sense of community (Maas, Van Dillen, Verheij, & Groenewegen, 2009), and, although dependent on neighborhood context, increased sense of safety (Maas, Spreeuwenberg, et al., 2009) and accessibility (Makris, 2015).

Young adults (ages 18-29) are an age group that can greatly benefit from the hypothesized effects of green space. Young adults are at a sensitive time in life when they are expected to cope with new stressors such as full-time work and parenthood (Bonnie, Stroud, Breiner, & others, 2015), while losing previous resources including school social environments and parental support. Given increased burdens and decreased available resources, it is perhaps not surprising that young adulthood is also when the majority of US adults transition from active to sedentary lifestyles (Neinstein, 2013). Concomitant increased responsibility and decreased resources can amplify socioeconomic and mental health inequalities, particularly among sensitive populations with a large preexisting burden:resource ratio (Bonnie et al., 2015).

Young adult Millennials (born approximately between 1982-2003, Bonnie et al., 2015) in American society face several physical, societal, and economic challenges that are greater compared to previous generations. Millennials are earning less compared to their parents at the same age (Chetty et al., 2016) and have greater levels of student loan debt, restricting societal roles such as homeownership and living independently from parents (Bleemer, Brown, Lee, & van Der Klaauw, 2014), despite desire by the majority of Millennials to do so (Xu, Johnson, Bartholomae, O’Neill, & Gutter, 2015). While the transition from active to sedentary lifestyles may not be exclusive to the Millennial generation, Millennials are, according to the Institute of Medicine (IOM), the first generation of the childhood obesity epidemic (Bonnie et al., 2015), exacerbating risks associated with transitioning to a sedentary lifestyle. In 2015 the IOM, in collaboration with the National Research Council, called for increased policy and research investment into the young adult population and identified challenges and next steps for young adult health-related research (Bonnie et al., 2015, pg.7):

*“The most immediate tasks are to improve data and research and to make a concerted effort to evaluate existing policies and programs at every level so as to achieve greater specificity and outcomes for young adults, while exploring new policies and programs.”*

Potential health benefits associated with green space can ameliorate several of the physical and psychological challenges prevalent in young adults. To leverage green space environments for improving young adult health, however, we must first understand how young adults value, perceive, and use green space, using methods that are conducive and likely to capture contextual characteristics of the young adult population.

Many methods have been used to assess green space perceptions, attitudes, and use, including interviews (Irvine, Warber, Devine-Wright, & Gaston, 2013), surveys (Gunnarsson, Knez, Hedblom, & Sang, 2016), and expert panels (Carrus et al., 2015). While valuable, these methods are difficult to replicate across large, diverse populations that are the cornerstone of environmental epidemiological studies. Recently, automated image processing of Google Street View imagery has been used to estimate amount of visible green space (Li et al., 2015) and identify characteristics such as barriers around, browning of, and vertical growth of green space that may influence green space perceptions (in review, data unpublished). While imagery may accurately capture visual characteristics of green space environments, contextual factors which influence greenspace impacts on health, such as perceived safety and use of greenspace, may be highly localized (Maas, Spreeuwenberg, et al., 2009) and temporally variant. Data sources must be considered which can capture inter-societal and temporal variations in attitudes, perceptions, and use.

Social media is a data source that has potential to capture characteristics and activity patterns of the young adult population. In the US, between 85% (Rainie, Fox, & Duggan, 2014) to 90% (Perrin, 2015) of young adults actively participate in social media and approximately 1/3 of young adults (Mitchell & Page, 2014) completely rely on social media as their source for news. One of the most active social media sites is Twitter, with 330 million active monthly users (<https://about.twitter.com/company>). Twitter posts are publicly searchable short messages or “Tweets” consisting of 140 characters or less. Recent research analyzing the content of tweets has provided insights into young adult sentiments and attitudes towards marijuana (Cavazos-Rehg et al., 2015a) and alcohol consumption (Cavazos-Rehg et al., 2015b). Twitter also has significant potential to capture temporal changes in attitudes, perceptions, and use. Twitter posts occur frequently, with more than 500 million Tweets daily (<https://about.twitter.com/company>). An analysis of Twitter posts focused on young adult sentiment and attitudes towards green space can advance our understanding of how green space impacts young adult health and how public health actors can leverage social media to increase positive green space impacts on health. However, a critical analysis of the current state of social media green space research, gaps in knowledge and needs among green space research experts, and the quality, quantity, and content of green space Twitter messages is required before a robust Twitter-based greenspace study can be performed.

**Specific Aims**

The purpose of this critical analysis (CA) is to evaluate the quality, quantity, and content in green-space related Twitter posts, and whether Twitter can be used in health-related green space research. The target population for the CA consisted of young adults living in or visiting Portland, OR. Portland has a population of 632,000, 10.3% of which are young adults (US Census Bureau, 2015). Portland is an ideal living environment for exploratory green space research because of significant concentration of and intra-urban heterogeneity in green space (Netusil et al., 2014). Specifically, this CA set out to answer the following questions:

1. What is the current state of research in using social media analytics (SMA) to evaluate green space use and perceptions?
2. What are the most common attitudes, perceptions, and uses mentioned in green space-related Tweets in the target area? How do they compare to previous studies that use alternate data sources and study designs?
3. What are the strengths, limitations, and recommended best practices in using Twitter for health-related green space research and/or health promotion/intervention?

**Methods**

*Literature Review*

To address specific aims one and two, I performed a literature review using PubMed, Scopus, Google Scholar, and PsycINFO search engines. Selection criteria included publication date between January 1, 2011 and December 1, 2016, English text, relevant to the research topic, and full-text availability, either through open-access publication or download via the George Washington University Himmelfarb Library, Oregon State University Valley Library, and/or Johns Hopkins Sheridan Libraries. Search terms for the literature review are shown below in Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| First Search Term | Modifier | Second Search Term | Modifier | Tertiary Search Term |
| green space | AND | social media |  |  |
| green space | AND | Twitter |  |  |
| park(s) | AND | social media |  |  |
| park(s) | AND | Twitter |  |  |
| natural environment(s) | AND | social media |  |  |
| natural environment(s) | AND | Twitter |  |  |
| green space | AND | young adult | AND | use |
| green space | AND | young adult | AND | sentiment |
| green space | AND | young adult | AND | attitude(s) |
| park(s) | AND | young adult | AND | use |
| park(s) | AND | young adult | AND | sentiment |
| park(s) | AND | young adult | AND | attitude(s) |
| natural environment(s) | AND | young adult | AND | use |
| natural environment(s) | AND | young adult | AND | sentiment |
| natural environment(s) | AND | young adult | AND | attitude(s) |

Table 1. Literature review search term combinations.

*Twitter Search*

To address specific aims two and three, green-space related Tweets were collected from December 8th 10am - December 10th 10am using Python v. 2.7 (van Rossum & Drake Jr, 2010) and the Python module Tweepy (Roesslein, 2015). Tweets were downloaded if they contained one or more keywords shown below in Table 2. In addition to Tweet content, the time of tweet and self-identified user location were downloaded. Tweets were then screened for self-identified locations within the Portland, OR area and Tweet content relevant to green space. The Python script created to download Tweets and convert to csv format is available at https://github.com/larkinandy/MPH-Culminating-Experience.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Twitter search keywords | | | | |
| backyard(s) | forest(s) | lawn(s) | pasture(s) | stream(s) |
| bush(es) | garden(s) | leaves | plant(s) | trail(s) |
| crop(s) | grass | mountain(s) | prairie(s) | tree(s) |
| field(s) | hay | nature | river(s) | woods |
| flower(s) | lake(s) | park(s) | riverside | yard(s) |

Table 2. Twitter search keywords.

*Tweet Analysis*

A thematic analysis was performed on Tweets that met selection criteria. Tweets were coded in NVivo (QSR International Pty Ltd. Version 11.4.1, 2016) into themes developed a priori based on the literature review and discussions with experts in green space research (P. Hystad, personal communication, January 2017). Each Tweet was also assigned a sentiment score of positive, negative, or neutral, based on the lead researcher’s impression of the sentiment the author of the Tweet wished to communicate with the Tweet. The quality, quantity, and context (coded themes and sentiment) of screened Tweets was then compared to previous studies.

*Automated Tweet Analysis*

Tweet coding by hand is impractical for large study areas or studies that include a time series component. To evaluate the potential of automated tweet analysis, the researchers utilized a Natural Language Parser (NLP) and Recurrent Neural Network (RNN) provided courtesy of the Stanford CoreNLP (<http://stanfordnlp.github.io/CoreNLP/index.html>). The NLP (Manning et al., 2014) identified verbs, adverbs, modifiers, adjectives, and nouns syntactically associated with Twitter search keywords, which were then used in a word frequency analysis (i.e. word clouds). If successful, limiting word frequency analysis to those syntactically linked with keywords will provide insight into uses and perceptions of greenspace. For example, frequent links of beautiful with trees suggests that tree components of green space contribute to aesthetically-oriented activities such as sight watching and positive green space attitudes. An example Tweet processed by the Stanford NLP is shown below in Figure 1.

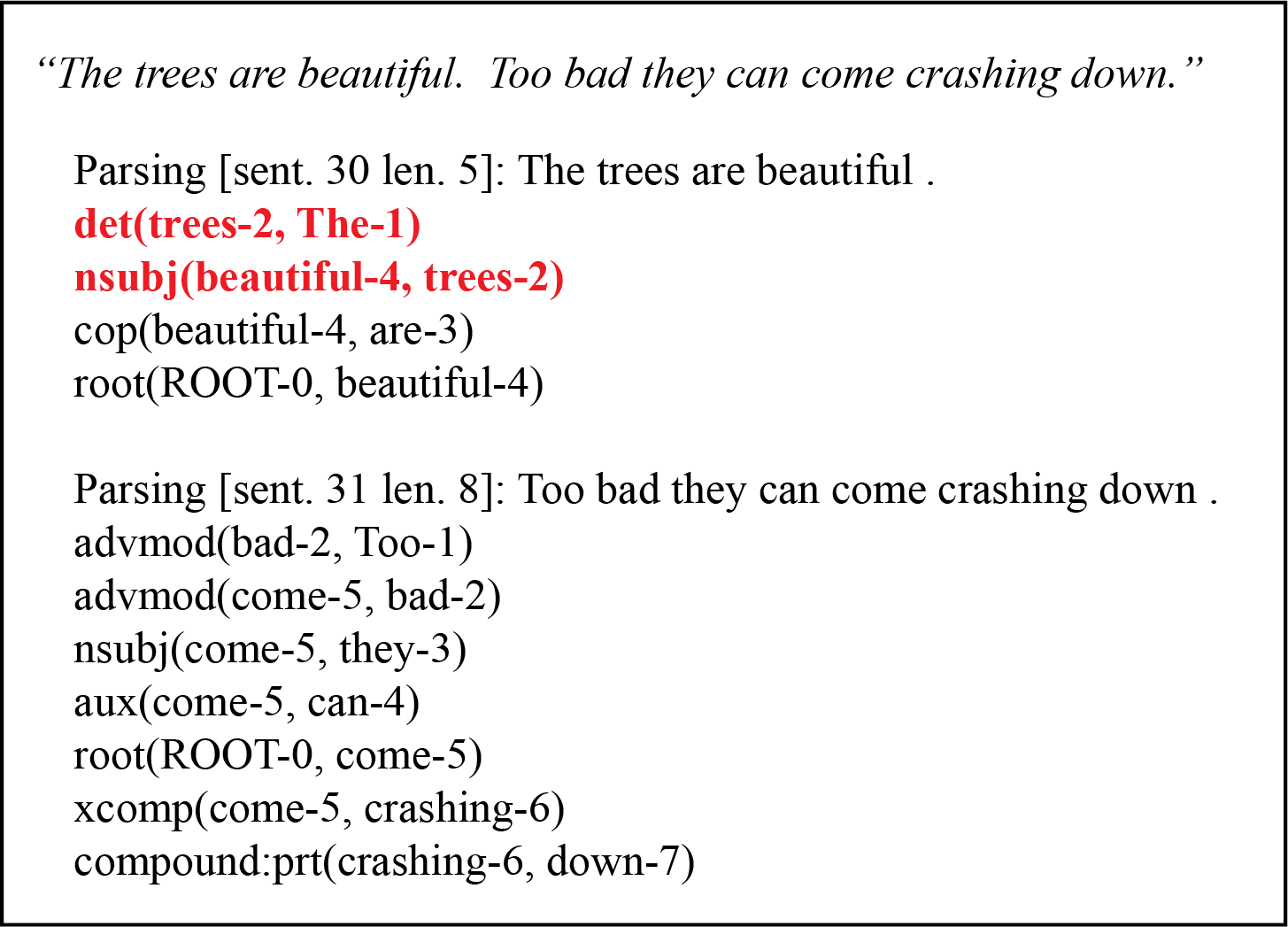


Figure 1. Example tweet and output produced by the Stanford NLP. Keyword syntactic relationships are highlighted in red.

The RNN was trained using the Stanford Sentiment Treebank (Socher et al., 2013) and used to predict sentiment (positive, negative, or neutral) of screened Tweets. Agreement between human and neural network sentiment scores were then compared for all three sentiment categories, and separate word clouds were generated for subsets of Tweets with positive and negative sentiment.

**Findings**

Eighteen papers were found during the literature search, 17 of which were accessible for review (Table 3). Papers directly relevant to green space research were found using Google Scholar and PubMed. Papers relevant to social media-based health interventions were found using PubMed and Scopus. One manuscript (Hand, Kenne, Wolfram, Abram, & Fleming, 2016) was added to the list based on a professional recommendation.

|  |  |  |
| --- | --- | --- |
| Authors | Title | Year |
| Afzalan & Muller. | The role of social media in green infrastructure planning: A case study of neighborhood participation in park siting | 2014 |
| Andkjær & Arvidsen. | Places for active outdoor recreation-a scoping review | 2015 |
| Falxa-Raymond, Svendsen, & Campbell, | From job training to green jobs: A case study of a young adult employment program centered on environmental restoration in New York City, USA | 2013 |
| Foo, Martin, Polsky, Wool, & Ziemer, | Social well-being and environmental governance in urban neighbourhoods in Boston, MA | 2014 |
| Hand, Kenne, Wolfram, Abram, & Fleming. | Assessing the Viability of Social Media for Disseminating Evidence-Based Nutrition Practice Guideline Through Content Analysis of Twitter Messaging | 2016 |
| Lin, Fuller, Bush, Gaston, & Shanahan. | Opportunity or Orientation? Who Uses Urban Parks and Why | 2014 |
| MacKerron & Mourato. | Happiness is greater in natural environments | 2013 |
| Madureira, Nunes, Oliveira, Cormier, & Madureira. | Urban residents' beliefs concerning green space benefits in four cities in France and Portugal | 2015 |
| Makris. | Separate, Different, but Not Isolated: How Youth in Public Housing Relate to Their Gentrified Community | 2015 |
| Palomino, Taylor, Göker, Isaacs, & Warber. | The online dissemination of nature-health concepts: lessons from sentiment metabolism perspective | 2016 |
| Park, Reber, & Chon. | Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement | 2016 |
| Pereira, Christian, Foster, Boruff, Bull, Knuiman, & Giles-Corti, | The association between neighborhood greeness and weight status: an observational study in Perth Western Australia | 2013 |
| Roemmich, Balantekin, & Beeler. | Park-Like Campus Settings and Physical Activity | 2014 |
| Sato & Conner. | The quality of time in nature: how fascination explains and enhances the relationship between nature experiences and daily effect | 2013 |
| Visser, Sichling, & Chaskin, | Hot times, hot places. Youths' risk perceptions and risk management in Chicago and Rotterdam | 2016 |
| Wood, Guerry, Silver, & Lacayo. | Using social media to quantify nature-based tourism and recreation | 2013 |

Table 3. Final papers included in analysis, ordered by alphabet..

*Question #1: What is the current state of research in using social media analytics (SMA) to evaluate green space use and perceptions?*

Social media is a relatively new data source for green space-related research. Among the literature review papers, three papers utilized social media for environmental research. In 2014, Quercia and associates (2014) utilized Flickr comments at key locations within Boston to estimate transportation routes that optimize beauty, quiet, or happiness rather than transportation time. Using a crowd-sourced based evaluation, participants rated routes optimized for beauty as more beautiful than shortest distance routes. Quercia’s paper demonstrated that social media comments can accurately capture latent constructs of the built environment such as beauty at specific geographical locations. However, the environmental composition of each rated location, including green space, was not considered.

In a similar research study, Wood and associates (2013) collected Flickr metadata and comments to estimate visitation rates at 836 recreational parks across the world. Estimates were then compared to recorded visitation rates. Results demonstrated that Flickr meta data and comments not only captured average visitation rates, but also repeated time series events, such as seasonality, and unique time series events, such as the marked decrease in tourism following September 11, 2001. However, Wood and associates also did not consider the composition of recreational parks, and generalizability is limited to tourism.

In 2016, Palomino and associates published a proof of principle study, demonstrating that Twitter can be combined with sentiment analysis to evaluate nature-related concepts. The authors collected 176,494 tweets containing hashtags related to nature deficit disorder (e.g. #lastchildinthewoods). Analysis of the collected data found that non-individual actors (agencies, non-profit groups, etc.) contributed to most of the Twitter feed, and that hashtags associated with positive sentiments were more likely to be retweeted than those with negative sentiment. Palomino and associates demonstrated that large scale green space-related research with social media is feasible. It is unclear, however, how or if tweet popularity can be used to infer green space perceptions and use among the participating Tweeters, particularly given the dominance of non-individual actors (e.g. social groups, businesses, non-profit agencies) in the sampled set of tweets. It should also be noted that hashtags are associated with significant self-selection bias, and concerns have been raised over the validity and generalizability of studies that use hashtag-based sampling (Kim, Huang, & Emery, 2016). Although accuracy of automated sentiment analysis of Tweets has been robust in previous studies (Meehan, Lunney, Curran, & McCaughey, 2013) sentiment accuracy was not evaluated in this particular study.

In summary, social media has largely remained unutilized in green space-related research. Publication dates of the three papers found range from 2014-2016, suggesting that social media might be on the verge of becoming more integrated in green space research. While these papers demonstrate the potential of social media for capturing latent constructs, time series, unique events, and sentiment of the built environment, none of the papers provide a clear framework for using social media to estimate green space perceptions, either in general or among the young adult population. Palomino and associates demonstrated the importance of identifying non-individual Twitter actors and frequent Tweeters, and potential positive attitudes associated with nature and/or bias due to hashtag-based sampling.

*Question #2: What are the most common attitudes, perceptions, and uses mentioned in green space-related Tweets in the target area? How do they compare to previous studies that use alternate data sources and study designs?*

Green space attitudes

Previous studies have shown that attitudes towards green space are highly localized and dependent on societal context. In comparing multiple cities, Madureira et al. (2015) found that green space values and beliefs about conferred green space benefits differed both between and within cities. In Hoboken, New York, Makris (2015) observed attitudes towards parks were dependent on access to parks, both in terms of amount of access and relative access compared to past years. In Boston, Massachusetts, Foo and associates (2015) found that benefits conferred from the urban environment are dependent on perceived stewardship, control, safety, and familiarity. In Chicago, Illinois and Hoboken, New York, perceived park safety among teenagers was dependent on the number (increased safety) and perceived threat (decreased safety) of people in the park (Makris, 2015; Visser, Sichling, & Chaskin, 2016). Together, these studies suggest that local context is important in determining perceived greenspace benefits.

After removing retweets and tweets irrelevant to green space, 602 unique Tweets from Portland, OR were collected and coded in NVivo. Several Tweets provided time-sensitive context of specific natural environment locations, with both positive (greater perceived benefit) and negative (lesser perceived benefit) sentiment about local green space.

*“20-year-old man found dead in Gresham park; homicide investigation underway.” - negative sentiment, safety context*

*“Trees in Normandale Park did not have a good time in ice storm.” - negative sentiment, safety and familiarity contexts*

*“We 're thrilled that WRDA passed last week with the Columbia River Basin Restoration Act to clean up toxics in the Columbia!” - positive sentiment, safety and control contexts*

*“the Elliot State Forest should stay in public hands.” - ambiguous (neutral) sentiment, control and accessibility context*

Personal attitudes towards greenspace is strongly associated with both amount of green space use and short-term psychological benefits. In Brisbane, Australia, Lin et al. (2014) found that, although both were significant, affinity for green space was more strongly associated with park usage than distance to nearest park. In New Zealand, fascination with nature acted as an effect modifier on increased positive affect following time spent near green space in New Zealand (Sato & Conner, 2013). Young adults who found green space-related employment in New York City reported increased affinity for green space post- compared to pre- employment (Falxa-Raymond, Svendsen, & Campbell, 2013).

Tweets with both positive and negative sentiment provided insights into individuals who had high affinity for natural environments. None of the Tweets, however, contained text indicative of low affinity for natural environments.

*“I grew up along the Hudson River. It used to be a toxic sewer. I took a nice swim in it recently. I 'm a-ok with EPA.” – positive sentiment, high affinity for natural environments*

*“Yeah, let's sell the National Parks to rich guys, let the poor look at pictures!” – negative sentiment, high affinity for natural environments*

*“I used to garden a lot when I was a kid, so I have a huge love for plants, even if I kill the more difficult ones.” – ambiguous sentiment, high affinity for natural environments*

User names are publicly available for all Twitter posts, allowing researchers to follow tweeters over time and consider personal affinity for green space in study results. Following tweeters over time can potentially capture long term attitudes and change in attitudes towards green space, which if successful can facilitate research efforts at evaluating chronic effects of cumulative green space exposure (P. Hystad, personal communication, January 2017). However, retaining personal information such as user names and estimating personal characteristics without informed consent requires ethical considerations that are normally not required when performing an ecologically based Twitter analysis.

Green space use

Previous research also suggests that young adult uses of green space are dependent on the geographical setting. Roemmich et al. (2015) found greater percentage of young adults performing physical activities at parks compared to college campuses.

Not surprisingly, a large percentage of Twitter posts described user activities/interactions with green space. Twitter posts about activities had both positive and negative sentiments and captured a wide variety of activities. In addition to commonly expected activities, such as physical activity and admiring beautiful landscapes, many unexpected uses of green space were tweeted.

*“I love running on this trail.” – positive sentiment, physical activity*

*“Frosty leaves! Beautiful.” – positive sentiment, aesthetic activity*

*“Just lay down to chill and a damn tree falls down on the house. Come on nature!” – negative sentiment, misc. activity*

*“Our ancient backyard play structure has transformed a lot of children from strangers into friends” – positive sentiment, social activity.*

*“@OleLatteCoffee's pine-flavored latte is made with Douglas Fir trees strait from the cart owners backyard.” – neutral sentiment, misc. activity.*

In addition to Tweets containing useful contextual information, a large number of Tweets had ambiguous sentiment and context. This was expected, as Twitter, either by design or evolution, is a media for short statements, conducive to abbreviations, slang, and acronyms that can be difficult to interpret.

*“Portland Snow Day = hibernation?” – unknown context, sentiment*

*“if plants eat meat, don't vegans cancel out? ain't that PEMDAS or something” – unknown context, sentiment*

*“Stone lika mountain? , wetter than a fountain?? , told em he go have to pay ?for this ? like its acountin…” – unknown context, sentiment.*

As shown in several Tweet examples within this section, several collected tweets described interactions between green space and Tweeters that were specific to snow storm events, a unique but not uncommon occurrence in Portland, Oregon. Tweeters described episodic beauty in frozen trees, leaves, and gardens, while also expressing fear and anxiety over falling tree branches. Most green space research takes place during peak greenness, typically late summer in the US. Here Twitter presents a unique opportunity to frequently capture green space attitudes, perceptions, and use throughout the year.

In agreement with previous studies suggesting hashtag-based sampling may be biased, only 42 out of 602 tweets (7.0%) contained one or more keywords (including proper noun hashtags such as #CentralPark) within part of all of the hashtag within the Tweet text. Sentiment scores for all 42 tweets were either positive or neutral. That’s not to say that hashtags were absent from hashtags with negative sentiment. Rather, negative sentiment hashtags did not directly relate to green space. Example negative sentiment hashtags include #Snowpocalypse, #snowmageddon, and #PDXTST.

Tweets related to green space for health promotion/prevention

Only one tweet within the collected dataset contained a health promotion/prevention message, encouraging physical exercise in green space environments. Messages that indirectly facilitate health promotion/prevention include updates regarding park access by the Portland Department of Parks and Recreation, and advertisements from subscribed twitter feeds.

*“Day 9 of the Crazy Wild Love Holiday Challenge is to take a walk. Get out in nature.” – health promotion message*

*“Due to ice/snow in Washington Park, the Garden will open at 10am on Fri, Dec 9.” – park access notification*

*“The latest DIY ~ Home And Gardening Daily!” - advertisement*

The lack of health promotion tweets may be a cause of snowy weather, where most Portland residents inexperienced with icy conditions are safer indoors rather than exercising in outdoor green space environments. Additional sampling during a non-episodic event and alternative seasons is needed to better estimate the frequency and content of health promotion green tweets.

Automated Twitter Analysis

Results of the automated Twitter sentiment analysis are shown below in Table 4. Automated and human sentiment scores are in agreement for 69%, 39%, and 61% of tweets classified as negative, neutral, and positive by human sentiment scoring, respectively. While the human scoring has an equal number of negative and positive sentiments, sentiment scoring using the automated RNN is biased towards negative sentiment. This may be due to domain mismatch, as the RNN was trained on movie reviews rather than twitter posts, and the overall level of grammar most likely differs. The most common divergence between machine and human sentiment scores occurred when one of the methods classified a statement as neutral. If we consider only a polar opposite sentiment as divergence between automated and human sentiment scoring (i.e. one method scores positive and the other negative), then human and automated sentiment scores are in agreement for 90% of tweets. It should also be noted that human sentiment was scored by a single researcher rather than the more common practice of using multiple scorers, which may add personal bias (towards either direction) to sentiment score results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| automated sentiment | human scored sentiment | | |  |
| negative | neutral | positive | total |
| negative | 62 | 106 | 26 | 194 |
| neutral | 11 | 88 | 11 | 110 |
| positive | 17 | 34 | 59 | 110 |
| total | 90 | 228 | 96 |  |

Table 4. Comparison of human-based and automated sentiment analysis.

Sarcastic Tweets are largely responsible for tweets in which human and automated sentiment scores are in opposite directions. While sarcastic tweets make up a small percentage of the Twitter dataset, they often correspond to strong emotional feelings that could be valuable for green space research.

*“Yeah, let's sell the National Parks to rich guys, let the poor look at pictures!” – positive automated and negative human sentiment score*

While the automated sentiment analysis may seem poor at first glance, many of the mismatches can be attributed to ambiguous sentiment in many Tweets.

*“Snow from the front.”*

*“3 Dams to Be Removed in American West to Restore Rivers.”*

*“It snowed in Portland”*

Tweets such as ‘it snowed in Portland’ can be considered positive, negative, or neutral, depending on background context missing from the Tweet. Given the large number of Tweets with ambiguous sentiment, a high level of agreement between automated and human scoring would suggest that the automated algorithm is overfitting predictions to bias results toward the personal perceptions of the human scorer rather than accurately capturing the original Tweeter sentiment. Larger evaluation sets with green space specific tweets and multiple human scorers are needed to better evaluate the potential of automated sentiment analysis.

Word clouds produced by the NLP syntactic frequency analysis are shown below in Figure 2. The top left word cloud corresponds to frequency analysis based on all tweets. Prevalent keywords include city, frozen, and fell. Compared to the thematic analysis performed in NVivo, the frequency based analysis provides little insight into societal and personal context underlying the dataset of tweets.



Figure 2. Word clouds from the automated syntactic tweet analysis. From top left moving clockwise: Frequency analysis based on all tweets. Top right: Frequency analyses based on a subset of tweets containing garden, flower, or plant keywords. Bottom right: a subset of tweets classified as negative by automated analysis. Bottom left: a subset of tweets classified as positive by automated analysis.

While frequency analysis of the entire tweet dataset provided little insight, guided analyses, either based on sentiment or thematic groups of keywords, proved more useful. In a subset of tweets containing keywords garden, flower, and/or plant (Figure 2, top right), prevalent words included unique nouns such as community, park, and snow, and adjectives including love, and beautiful. Although the subset analysis still provides less context compared to human coding, more context was provided than the unfiltered analysis. Similarly, stratifying tweets by automated positive and negative sentiment classifications (Figure 2, bottom left and right, respectively) identified beautiful and frozen as keywords of positive expressed sentiments, and fell, hood, and city as keywords in negative expressed sentiments. This agrees with the human thematic analysis, which found most positive sentiments related to aesthetic beauty of frozen green space objects, and most negative sentiments related to the danger of falling tree branches and being stuck on or not able to access Mt. Hood due to road closures.

In summary, automated Twitter analysis has potential to analyze large volumes of tweets. In this CA, 602 tweets were collected from a city with a population of 602,000 within 48 hours in the winter, a tweet density of one green space tweet per 2000 people per day. Assuming similar tweet densities for other areas and seasons (most likely a conservative estimate given green space use increases during warmer seasons) collecting tweets for New York City over the course of a year would net 1.46 million green space tweets. As demonstrated in this CA, episodic events can strongly influence acute local perceptions and use of green space, limiting generalizability of fixed time sampling. Unfortunately, unfiltered automated analysis yielded little useful information. Automated analyses of subsets of tweets based on a priori search themes (tweets containing related keywords such as flowers, garden and plant) and sentiment fared better, and suggests that automated analysis can be best utilized in a mixed methods approach, using a priori hypotheses to drive subset analyses. For example, a preliminary random sample of Tweets can be coded in NVivo to identify keywords that capture Foo and associates’ factors which drive perceived benefits from urban environments (stewardship, political and economic control, safety, and familiarity and social control), which can then be used to subset and guide an automated Twitter analysis.

*What are the strengths, limitations, and recommended best practices in using Twitter for green space-related research and health promotion/intervention?*

Green space research

In this CA, Tweets that were collected over 48 hours from Portland, OR provided contextual information relevant to green space attitudes, perceptions, and use. Previous studies also demonstrated utility in estimating latent constructs such as beauty and happiness and use such as visitation rates from social media.

However, there are several factors which limit social media generalizability, particularly regarding inference to a specific demographic population such as young adults. Firstly, many social media platforms including Twitter do not provide demographic information about users. While machine learning methods have been developed to estimate demographic profiles for Tweeters based on meta data (Burger, Henderson, Kim, & Zarrella, 2011; Culotta, Kumar, & Cutler, 2015), assigning demographic labels to users that haven’t consented to share personal demographic information is ethically questionable and potentially unreliable. Social media analysts therefore need to consider the possibility of ecological fallacy, in which the overall perceptions and use of green space described in Twitter may not be the same as a population group of interest, such as young adults. Secondly, while 85-90% of young adults in the US actively use social media, electronic resources required to use social media are inequitably distributed, leading to underrepresentation of underserved populations in social media activity (Sloan, Morgan, Burnap, & Williams, 2015). Social media preferences and levels of use also significantly differ between and within groups, and continually shifts as new social media platforms become available and younger generations seek alternative social media sties (Madden et al., 2013). Utilizing multiple social media data streams will increase the likelihood that social media analyses capture the perceptions and uses of multiple personality types and online communities.

Hashtags in the collected sample of tweets were strongly biased towards positive sentiment and represented a small percentage of green space-related Twitter communications. Unless there is compelling evidence to do so, hashtag-based sampling is not recommended. However, hashtags contain important information, as they are indicators of topics that are important to the tweeter, and can be used to identify characteristics of those passionate about green space which can be contrasted to the general Twitter population.

Green space health promotion/intervention

Although the tweet analysis yielded little information directly relevant to health promotion/intervention, we can extrapolate best practices from social media health promotion programs targeting other health behaviors/outcomes described in literature review papers. One of the most common used measures for estimating the popularity of reach of a tweet is the number of times the tweet is forwarded by a follower on their own Twitter stream, known as a ‘retweet’. In 2016, Hand and associates found that Twitter followers of hashtags related to heart failure were more likely to retweet messages from official organizations such as government agencies, charities and patient advocacy groups. In an analysis of tweets from the American Diabetes Association, American Heart Association, and American Cancer Society, Park, Reber and Chon (2016) found that personal messages encouraging action were most often retweeted by followers. In the Portland, OR community, there are multiple active Twitter groups that encourage green space activities, including @BikePortland, @NWTrailRuns, @PDXParksRec, and @PDXForestPark, with follower sizes ranging from 824 to 23,400. Given the highly localized context of green space impacts on health, encouraging green space related health promotion messages through already existing local groups such as these provides opportunities to tweet messages that meet recommendations from literature review findings and relevant to green space opportunities in the Portland community. Example public health intervention/promotion tweets include notices about community group activities in local parks, and opportunities to get involved in community gardens. It should be noted that these types of tweets are likely already part of the Portland Twitter community, and green space-related Twitter intervention efforts may be better focused on recruiting the population of interest to follow, participate, and retain connections to Twitter communities rather than to change the Twitter messages. Additional Twitter sampling and analyses during seasons with traditionally greater green space related activities are needed.

**Conclusions**

Green space-related tweets have potential to capture latent constructs and localized context which mediate green space impacts on human health. Real time tweet collection and analysis presents opportunities to capture unique, episodic beneficial and harmful green space interactions. Automated sentiment analyses include significant misclassification, but the number of polar opposite (negative, positive) misclassifications are small. Future research is required before automated thematic analyses are possible. Next steps for automated thematic analyses include a priori development of a list of keywords for each thematic group or latent construct of interest. Tweets with green-space related hashtags comprise a small percent of all green-space related tweets and biased towards positive sentiment. Hashtag-based sampling is not recommended. Local groups are recommended for a Twitter mediated green space health promotion/intervention program. A comprehensive sampling of Tweets from multiple seasons is needed, however, to determine if intervention efforts are best invested in improving Twitter messages or in encouraging people to follow, participate, and stay in the green space Twitter community.

**Acknowledgements**

I would like to acknowledge Professor Perry Hystad for lending his experience in green space related research to this CA. I am also grateful to Twitter and Tweepy developers for providing open source social media tools to non-profit researchers. Finally, thank you to Professor Christina Heminger for providing guidance and mentorship throughout the project.

**References**

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo*. Sage Publications Limited.

Beyer, K. M., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F. J., & Malecki, K. M. (2014). Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. *International Journal of Environmental Research and Public Health*, *11*(3), 3453–3472.

Bleemer, Z., Brown, M., Lee, D., & van Der Klaauw, W. (2014). Debt, jobs, or housing: what’s keeping millennials at home? Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2530691

Bonnie, R. J., Stroud, C., Breiner, H., & others. (2015). *Investing in the health and well-being of young adults*. National Academies Press. Retrieved from <https://www.nap.edu/catalog/18869/investing-in-the-health-and-well-being-of-young-adults>

Bratman, G. N., Hamilton, J. P., & Daily, G. C. (2012). The impacts of nature experience on human cognitive function and mental health. *Annals of the New York Academy of Sciences*, *1249*(1), 118–136.

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1301–1309). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2145568

Carrus, G., Scopelliti, M., Lafortezza, R., Colangelo, G., Ferrini, F., Salbitano, F., … Sanesi, G. (2015). Go greener, feel better? The positive effects of biodiversity on the well-being of individuals visiting urban and peri-urban green areas. *Landscape and Urban Planning*, *134*, 221–228.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., & Narang, J. (2016). *The Fading American Dream: Trends in Absolute Income Mobility Since 1940*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22910

Culotta, A., Kumar, N. R., & Cutler, J. (2015). Predicting the Demographics of Twitter Users from Website Traffic Data. In *AAAI* (pp. 72–78). Retrieved from https://pdfs.semanticscholar.org/b8f8/980efb2684ba2a0b7ddf8ea88a9fe09f2ade.pdf

Ebisu, K., Holford, T. R., & Bell, M. L. (2016). Association between greenness, urbanicity, and birth weight. *Science of The Total Environment*, *542*, 750–756.

Falxa-Raymond, N., Svendsen, E., & Campbell, L. K. (2013). From job training to green jobs: A case study of a young adult employment program centered on environmental restoration in New York City, USA. *Urban Forestry & Urban Greening*, *12*(3), 287–295.

Foo, K., Martin, D., Polsky, C., Wool, C., & Ziemer, M. (2015). Social well-being and environmental governance in urban neighbourhoods in Boston, MA. *The Geographical Journal*, *181*(2), 138–146.

Gunnarsson, B., Knez, I., Hedblom, M., & Sang, \AA Ode. (2016). Effects of biodiversity and environment-related attitude on perception of urban green space. *Urban Ecosystems*, 1–13.

Hand, R. K., Kenne, D., Wolfram, T. M., Abram, J. K., & Fleming, M. (2016). Assessing the Viability of Social Media for Disseminating Evidence-Based Nutrition Practice Guideline Through Content Analysis of Twitter Messages and Health Professional Interviews: An Observational Study. *Journal of Medical Internet Research*, *18*(11). Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5128725/>

Hystad, P. (2017, January). Personal Interview.

Irvine, K. N., Warber, S. L., Devine-Wright, P., & Gaston, K. J. (2013). Understanding urban green space as a health resource: A qualitative comparison of visit motivation and derived effects among park users in Sheffield, UK. *International Journal of Environmental Research and Public Health*, *10*(1), 417–442.

Kim, Y., Huang, J., & Emery, S. (2016). Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *Journal of Medical Internet Research*, *18*(2). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4788740/

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, *14*(3), 675–685.

Lin, B. B., Fuller, R. A., Bush, R., Gaston, K. J., & Shanahan, D. F. (2014). Opportunity or orientation? Who uses urban parks and why. *PLoS One*, *9*(1), e87422.

Lovasi, G. S., Schwartz-Soicher, O., Quinn, J. W., Berger, D. K., Neckerman, K. M., Jaslow, R., … Rundle, A. (2013). Neighborhood safety and green space as predictors of obesity among preschool children from low-income families in New York City. *Preventive Medicine*, *57*(3), 189–193.

Maas, J., Spreeuwenberg, P., Van Winsum-Westra, M., Verheij, R. A., Vries, S., & Groenewegen, P. P. (2009). Is green space in the living environment associated with people’s feelings of social safety? *Environment and Planning A*, *41*(7), 1763–1777.

Maas, J., Van Dillen, S. M., Verheij, R. A., & Groenewegen, P. P. (2009). Social contacts as a possible mechanism behind the relation between green space and health. *Health & Place*, *15*(2), 586–595.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, social media, and privacy. *Pew Research Center*, *21*, 2–86.

Madureira, H., Nunes, F., Oliveira, J. V., Cormier, L., & Madureira, T. (2015). Urban residents’ beliefs concerning green space benefits in four cities in France and Portugal. *Urban Forestry & Urban Greening*, *14*(1), 56–64.

Makris, M. V. (2015). Separate, Different, but Not Isolated: How Youth in Public Housing Relate to Their Gentrified Community. In *Public Housing and School Choice in a Gentrified City* (pp. 171–189). Springer. Retrieved from http://link.springer.com/chapter/10.1057/9781137412386\_7

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)* (pp. 55–60). Retrieved from http://www.aclweb.org/website/old\_anthology/P/P14/P14-5.pdf#page=67

Meehan, K., Lunney, T., Curran, K., & McCaughey, A. (2013). Context-aware intelligent recommendation system for tourism. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on* (pp. 328–331). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6529508/

Mitchell, A., & Page, D. (2014). State of the news media 2014. *Pew Research Journalism Project*. Retrieved from http://radiogood.ru/goto/http://www.journalism.org/2014/03/26/state-of-the-news-media-2014-overview/

Neinstein, L. S. (2013). The new adolescents: An analysis of health conditions, behaviors, risks, and access to services among emerging young adults. *Professor of Pediatrics and Medicine, University of Southern California*. Retrieved from <https://pdfs.semanticscholar.org/4161/045a0760f501d9f155850510c368c469bc19.pdf>Palomino, M., Taylor, T., Göker, A., Isaacs, J., & Warber, S. (2016). The Online Dissemination of Nature–Health Concepts: Lessons from Sentiment Analysis of Social Media Relating to “Nature-Deficit Disorder.” *International Journal of Environmental Research and Public Health*, *13*(1), 142.

Park, H., Reber, B. H., & Chon, M.-G. (2016). Tweeting as Health Communication: Health Organizations’ Use of Twitter for Health Promotion and Public Engagement. *Journal of Health Communication*, *21*(2), 188–198.

Perrin, A. (2015). Social media usage. *Pew Research Center*. Retrieved from http://www.gbcinteractive.com/s/PI\_2015-10-08\_Social-Networking-Usage-2005-2015\_FINAL.pdf

Quercia, D., Schifanella, R., & Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 116–125). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2631799

Rainie, L., Fox, S., & Duggan, M. (2014). The web at 25 in the US. *Pew Research Center’s Internet & American Life Project*.

Roemmich, J. N., Balantekin, K. N., & Beeler, J. E. (2015). Park-like campus settings and physical activity. *Journal of American College Health*, *63*(1), 68–72.

Roesslein, J. (2015). Tweepy. *Python Programming Language Module*.

Sato, I., & Conner, T. S. (2013). The quality of time in nature: How fascination explains and enhances the relationship between nature experiences and daily affect. *Ecopsychology*, *5*(3), 197–204.

Shanahan, D. F., Bush, R., Gaston, K. J., Lin, B. B., Dean, J., Barber, E., & Fuller, R. A. (2016). Health benefits from nature experiences depend on dose. *Scientific Reports*, *6*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4917833/

Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PloS One*, *10*(3), e0115545.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., … others. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (Vol. 1631, p. 1642). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.383.1327&rep=rep1&type=pdf

Takano, T., Nakamura, K., & Watanabe, M. (2002). Urban residential environments and senior citizens’ longevity in megacity areas: the importance of walkable green spaces. *Journal of Epidemiology and Community Health*, *56*(12), 913–918.

van den Berg, A. E., Maas, J., Verheij, R. A., & Groenewegen, P. P. (2010). Green space as a buffer between stressful life events and health. *Social Science & Medicine*, *70*(8), 1203–1210. https://doi.org/10.1016/j.socscimed.2010.01.002

van Rossum, G., & Drake Jr, F. L. (2010). Extending and embedding Python, Release 2.7. *Python Software Foundation, Wolfeboro Falls*.Visser, K., Sichling, F., & Chaskin, R. J. (2016). Hot times, hot places. Youths’ risk perceptions and risk management in Chicago and Rotterdam. *Journal of Youth Studies*, 1–17.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*, 2976.

Xu, Y., Johnson, C., Bartholomae, S., O’Neill, B., & Gutter, M. S. (2015). Homeownership among millennials: The deferred American dream? *Family and Consumer Sciences Research Journal*, *44*(2), 201–212.