Towards Using Social Media as a Geospatial Tool for Measuring Design Impact on Human Experiences

David Tulloch[1], Wansoo Im[2]

[1]Rutgers University, New Jersey/USA · tulloch@crssa.rutgers.edu
[2]Medharry Medical College, Tennessee/USA

Abstract: This paper presents preliminary findings showing how Twitter usage in public parks and adjacent areas can help measure the impact that the park and its design has on the user experience. As an initial test of the potential for these applications, the paper compares Twitter user updates in two park areas in New York City: Prospect Park and Central Park. The results show that Twitter is used differently in the parks than it is near them. Results also reveal differences between the two landmark parks. The paper finishes with an exploration for ways that Twitter can be used as a meaningful tool for measuring ways that park design impacts the visitor experience in the park.

Keywords: Twitter, geotagged, parks, social media

1 Introduction

Since its launch in 2006, Twitter has grown into a global phenomenon. Broad evidence of Twitter as a serious tool came through its highly publicized role in the 2011 Arab Spring movement (BRUNS et al. 2013). Twitter has become an important source of data for academic research. ZIMMER & PROFERES (2013) conducted a systematic analysis of Twitter-based research and found it published in fields as diverse as medicine, sport sciences, economics, and physics. The most common approach to analysis within the 380 published studies was the use of content analysis.

The ubiquity of social media applications on mobile devices has created digital environments through which personal thoughts and feelings are frequently shared in very public outlets in real-time (PAUL & DREDZE 2011). Since many of those personal devices integrate geospatial technologies (like GPS) to locate the device and its user, the social media postings are also shared with locational data as an attribute. Admittedly, much of the content of social media postings can require a sophisticated knowledge of the context and contemporary language (even emojis), some basic analyses can be employed to identify larger patterns in messages (e. g. BOLLEN et al. 2010, KOULOUMPIS et al. 2011).

There are several different characteristics of Twitter that would make it a particularly useful tool for studying landscape architecture, if it proves to be reliable. While this study is using a fixed temporal window, Twitter supports analysis of data as events happen permitting timely applications. As a global technology, it is potentially universal in its applicability to quite distant and distinct landscapes. Access to the publicly accessible stream of data, either through Twitter API or Twitter Firehose, makes it possible for researchers to study distant places without travel. Twitter is also commonly described as a widely accessible application, which creates the possibility of reaching populations that are less accessible through other avenues.

Journal of Digital Landscape Architecture, 3-2018, pp. 227-233. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-642-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537642024.

We propose three conceptual tests for establishing whether Twitter user updates − tweets − reflect on the human experience of landscape architecture. These conceptual tests are:

1. Do tweets in a designed landscape change sufficiently in location and content to validate additional spatial analysis at the site scale?
2. Do tweets suggest a different focus or emotional state in the parks than nearby?
3. Can tweets be used to compare experiences in similarly designed landscapes?

With a modest dataset, an initial test can be conducted for each of the three conceptual tests. These can help establish whether Twitter updates − and to a lesser extent, big data sources − have functional utility for the scale of spatial analysis necessary for these parks.

2 Methods

For this study, Central Park and Prospect Park were identified as sites that exemplify parks and park experiences within a city, experienced substantial levels of visitors using social media, and have reasonably well-defined boundaries. For comparison purposes, the parks' adjacent neighborhoods were used, having the same weather and conditions, treating areas within 1 mile (roughly a 20 minute walk) as a basis for comparison (Figure 1).

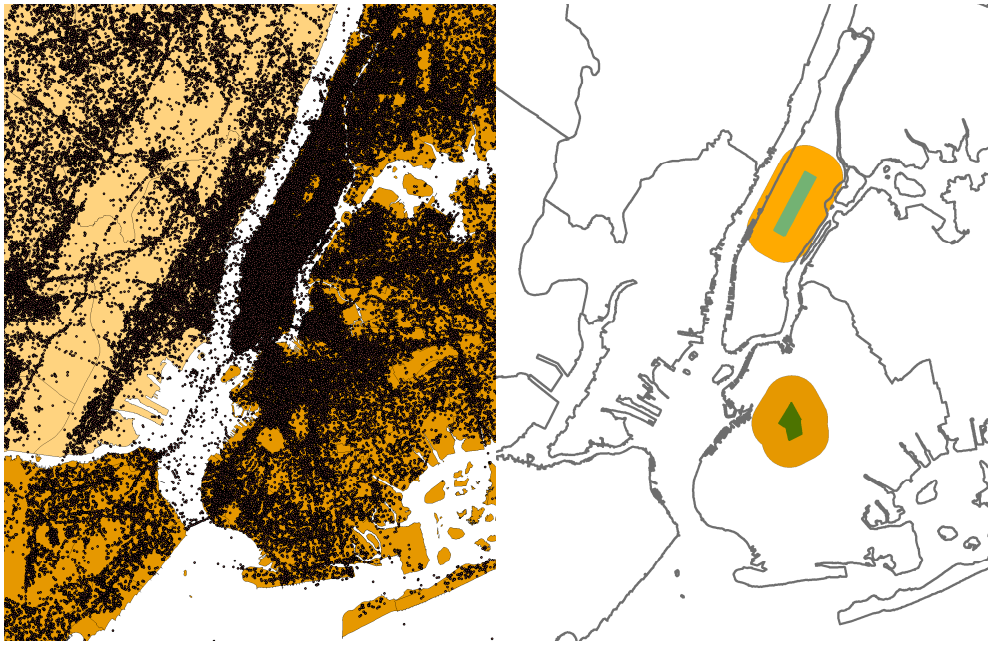

Fig. 1: When mapped, the Twitter landscape of New York City is an overwhelming mass of points (left). This paper isolates four zones from which tweets are extracted from: Central Park, near Central Park, Prospect Park and near Prospect Park.

A set of geocoded user updates in New York City were collected for a week in April 2015. Over that time period, the dataset for the study area included a total of 2,250,810 geocoded user updates, or tweets, identified as English in the tweet's metadata.

For the purposes of this project, four distinct zones were identified for the study period using the geolocated coordinates associated with each (the sets are mutually exclusive):

1. within Central Park (n = 88,471),

2. within 1 mile of Central Park but not inside Central Park (n = 1,693,975)

3. within Prospect Park (n = 11,730), and

4. within 1 mile of Prospect Park but not inside Prospect Park (n = 456,624).

3 Findings

3.1 Conceptual Test 1

If parks are experientially distinct zones within the city, then social media use would be expected to display similarly distinct patterns. A basic test of this hypothesis is the use of words that are linked directly to objects that are present in the park and not in the surrounding neighborhoods.

As an initial test, the number of geolocated tweets using the character string "Park" was counted and compared with the total number of tweets in that category. During the study period, 36 % of the geolocated tweets that were recorded within Central Park (zone 1) used the word park while 1.3 % of the geolocated tweets near the park used it. Over the same period, Prospect Park had an even higher usage rate of 44.6 % for the word "park" and the areas around it saw it used in 0.92 % of the tweets on the same day. Both parks experienced a rate that was more than 25 times higher than in the zones around them. But this example is problematic since some apps help users automatically tag the tweet with the place name.

A more useful test comes from the use of park object words which are much more prevalent in the park than the neighborhoods. For example, the word 'meadow' was used 50 times more inside the parks than in the neighborhoods around them (Table 1). Other park-specific object words – softball, fountain, and boat − were each mentioned 10 times more often inside the parks than in the neighborhoods. Additional location-specific words are shown in Table 1. The content of tweets inside the parks can be distinguished from those nearby, which supports a closer look at the second conceptual test.

Table 1: Comparison of use of words in geotagged tweets in parks (Central Park and Prospect Park) and near parks (within 1 mile of Central Park and Prospect Park). Comparison showing terms linked with specific features in the park.

	In Parks	Near Parks	Ratio (Parks/Near)
Reservoir	0.15 %	0.00 %	64.38
Meadow	0.14 %	0.00 %	52.54
Prospect	0.91 %	0.02 %	39.66
Central	8.65 %	0.26 %	33.55
Fountain	0.20 %	0.02 %	11.88
Boat	0.63 %	0.05 %	11.84
Softball	0.12 %	0.01 %	13.31
Lawn	0.11 %	0.01 %	9.98
Castle	0.09 %	0.01 %	5.97
Lake	0.35 %	0.11 %	3.29
Arch	0.93 %	0.34 %	2.77
Bridge	0.17 %	0.06 %	2.70
Pond	0.14 %	0.06 %	2.22
Hill	0.79 %	0.46 %	1.73

3.2 Conceptual Test 2

Unlike the specific features inside the parks (e. g., Belvedere Castle, Long Meadow), some park-like features and experiences also occur outside the park, but Twitter users encounter them less frequently outside the park. Visitors to the area can see a leaf, squirrel, bench or bird either in the parks or in the nearby neighborhoods. Analysis reflects higher rates of tweeting about these features within the parks than in the nearby neighborhoods (Table 2). Analysis found this much higher rate of usage for words including: Leaf, Squirrel, Water, Lawn, Landscape, Field, Bird, Sun, Sunshine, Cloud, Sky, Weather, Bench, and Grass. Each of these words were used within the parks at rates that were at least double those outside the parks. A possible reason for the dominant use of these words within the parks is that parks simply allow visitors to relax and notice every day features, like sunshine or urban wildlife. Another is that visitors to parks are conditioned to think of these features as part of the park experience and thus find that tweets about the parks need to include such words.

A notable opportunity for studying park experiences is the impact that the park has on the emotional state of its visitors. A popular technique of analysis of twitter updates has emerged in the form of sentiment analysis in which questions about the emotional state of Twitter users is examined through their use of key words and emoticons (KOULOUMPIS et al. 2011). For instance, the word happy was used more in the parks than in the areas around them. In Central Park it was used at a rate that was 40% higher than in the neighborhoods around it. In Prospect Park it was used at a rate that was 80% higher than in the neighborhoods around it. In contrast, the word stress was used twice as often in the combined neighborhoods as in the parks.

Table 2: Comparison of use of object words in geotagged tweets in parks (Central Park and Prospect Park) and near parks (within 1 mile of Central Park and Prospect Park). Comparison shows terms referencing features or objects that could be seen in or near the park.

	In Parks	Near Parks	Ratio (Parks/Near)
Lawn	0.11 %	0.01 %	9.98
Squirrel	0.07 %	0.01 %	8.73
Field	0.82 %	0.14 %	5.94
Landscape	0.04 %	0.01 %	5.47
Grass	0.14 %	0.03 %	4.70
Bench	0.10 %	0.02 %	4.12
Sun	1.68 %	0.51 %	3.27
Sunshine	0.19 %	0.06 %	3.05
Bird	0.26 %	0.10 %	2.59
Sky	0.60 %	0.26 %	2.33
Cloud	0.16 %	0.07 %	2.25
Water	0.44 %	0.21 %	2.09
Weather	0.44 %	0.21 %	2.08
Leaf	0.02 %	0.01 %	2.04
Crow	0.26 %	0.14 %	1.86
Tree	1.53 %	0.88 %	1.73
Leaves	0.05 %	0.03 %	1.56
Path	0.10 %	0.07 %	1.46
Shade	0.07 %	0.05 %	1.23
Rain	1.17 %	0.97 %	1.20

3.3 Conceptual Test 3

Comparing parks is potentially the most interesting application. As a simple measure, we can begin with basic place names as an extreme baseline proving that there is a distinction. Unsurprisingly, Twitter users in Central Park used the word “prospect” in 0.005 % of their tweets compared with 7.73 % of the tweets in Prospect Park. Conversely, Twitter users in Central Park used the work “central” in 9.79 % of their tweets compared with 0.034 % of the tweets in Prospect Park. The word “Manhattan” was used in 3.634 % of all tweets in Central Park but only 0.167 % in Prospect Park. The word “Brooklyn” saw a similar pattern, 8.981 % in Prospect Park but only 0.059 % in Central Park. Although less common overall, the word “island” was also used at a much higher rate in Prospect Park (0.108 % to 0.037 %), perhaps because Brooklyn is on or adjacent to Long Island. These results help establish that, indeed, social media users employ different words in these two landmark parks.

Both parks have meadows: Central Park has the Sheep Meadow while Prospect Park has the Long Meadow. The higher use of the term meadow in Central Park (0.16 %) than Prospect Park (0.02 %) may suggest a difference in the relative importance of each for visitors to the park or maybe a familiarity with the proper names. Both stand out dramatically compared to the use of meadow outside the park (0.003 %). However, specific features may be less useful or interesting than emotions or experiences.

Park-to-park comparisons of user content provide for comparisons of park design and user experiences and expectations. During the period of study, users in Prospect Park posted at a substantially higher rate about the words bird, path, play, fun and sun (Table 3). Over that same period, Central Park visitors used the words bench, beautiful, calm, land, tree and happy at substantially higher rates. The words leaves, ball, run, and cloud are examples of words used at roughly the same rate in each park.

Some of these differences might simply reflect local naming conventions. But it is worth exploring why grass is mentioned more in Prospect Park while the word beautiful is used more in Central Park. A more calibrated version of this tool might facilitate a relatively deep analysis of individual responses to the types of spaces presented by each park.

Table 3: Comparison of tweets in Central Park with those in Prospect Park

	Central	Prospect	Ratio (C/P)
Central	9.79 %	0.03 %	287.18
Meadow	0.16 %	0.02 %	9.28
Field	0.91 %	0.14 %	6.29
Squirrel	0.08 %	0.03 %	3.14
Landscape	0.04 %	0.02 %	2.52
Bench	0.11 %	0.04 %	2.52
Fountain	0.21 %	0.09 %	2.23
Bridge	0.18 %	0.09 %	1.95
Lawn	0.11 %	0.06 %	1.87
Sky	0.63 %	0.36 %	1.75
Pond	0.15 %	0.09 %	1.60
Beautiful	1.76 %	1.22 %	1.44
Weather	0.45 %	0.36 %	1.26
Land	0.73 %	0.58 %	1.26
Wall	0.10 %	0.09 %	1.22
Tree	1.56 %	1.32 %	1.18
Rock	0.44 %	0.41 %	1.08
Cloud	0.16 %	0.15 %	1.07
Pine	0.14 %	0.14 %	0.96
Ball	0.72 %	0.77 %	0.94

Table 3 (continued)

	Central	Prospect	Ratio (C/P)
Leaves	0.05 %	0.05 %	0.93
Sunshine	0.18 %	0.20 %	0.89
Hill	0.78 %	0.89 %	0.88
Sun	1.65 %	1.93 %	0.85
Play	0.98 %	1.31 %	0.75
Path	0.10 %	0.14 %	0.73
Grass	0.14 %	0.20 %	0.70
Music	0.64 %	1.12 %	0.57
Bird	0.24 %	0.43 %	0.56
Rain	1.06 %	1.97 %	0.54
Lake	0.31 %	0.64 %	0.48
Water	0.35 %	1.07 %	0.33
Prospect	0.00 %	7.73 %	0.00

4 Moving Forward

While CAD, Photoshop and GIS may remain the most commonly used digital tools for many landscape architects, these results show that social media might someday be an additional tool for some. Much needs to be done before this can become a practical application in a professional office. But the initial results reported in this paper demonstrate a level of utility that demands continued development. Furthermore, this may prove a useful parallel to other big data applications in design.

An immediate priority for this research is to develop a more expansive lexicon of appropriate searchable terms to expand these findings. Like much Twitter-based content analysis and sentiment analysis that has been conducted in other fields, the expanded vocabulary becomes a central tool and a valuable point of reference for future research. If the parks continue to produce consistent results, the methods could be used to ask a variety of other questions. Does Twitter differentiation regarding parks remain consistent across different seasons? Do other parks perform the same way? Comparisons across climate zones and geographic regions might reveal some fundamental differences in perspectives.

Neighborhoods with more street trees, or more generous sidewalks might be expected to have results that, to a lesser degree, echo the results in park. Comparing different neighborhoods with an index of park-ness is a logical step, once the data are more clearly understood. What other parts of the city have measurable degrees of “park-ness”? Which neighborhoods most need green infrastructural investments from the city, to improve the quality of the urban experience?

Were this developed to the point of serving as a calibrated tool, larger comparisons of park experiences could be tested. Which types of parks produce greater happiness or more acute

attention to details? Could park experiences be compared with the designers expressed goals? Could tweets produce a strong measure of physical activity and serve as a tool for monitoring human health?

As computer scientists and software engineers expand the exploration of social media tools, there remain technical opportunities for developing more advanced parsing or computer learning approaches as well as integrating more specific forms of sentiment analysis including the interpretation and classification of hashtags and emoticons as well as an expanded landscape lexicon. The rapidly growing research literature supporting geospatial analysis of Twitter (e. g. WALTHER & KAISSER 2012, MITCHELL et al. 2013.) holds intriguing possibilities for moving in these directions. But for our purposes, the digital landscape architecture community needs to focus these tools on designed landscapes in order to reap benefits for designed landscapes.

5 Acknowledgements

A special thanks is due to Dr. Kyumin Lee and Dr. Kyoungwon Suh whose help in accessing the data was essential to the project. Additional thanks are due to Samantha Vickers-Hymowitz and Jason Mallonga, Aresty Undergradaute Research Fellows whose questions about the parks and Twitter data helped focus an otherwise enormous exploration.

References

BOLLEN, J., ALBERTO, P. & MAO, H. (2010), Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Proceedings of International Conference on Web and Social Media, 450-453.

BRUNS, A., HIGHFIELD, T. & BURGESS, J. (2013), The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. American Behavioral Scientist, 57 (7), 871-89.

KOULOUMPIS, E., WILSON, T. & MOORE, J. (2011), Twitter sentiment analysis: The good, the bad, and the OMG! Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 538-541.

MITCHELL, L., FRANK, M., HARRIS, K. D., DOBBS, P. S. & DANFORTH, C. M. (2013), The geography of happiness: Connecting Twitter sentiment and expression, demographics and objective characteristics of place. PLOS One, (May 29), p. e64417.

PAUL, M. J. & DREDZE, M. (2011), You are what you tweet: Analyzing Twitter for Public Health. Proceedings of International Conference on Web and Social Media, 265-272.

WALTHER, M. & KAISSER, M. (2012), Geo-spatial event detection in the Twitter stream. Advances in Information Retrieval, Springer.

ZIMMER, M. & PROFERES, N. J. (2013), A topology of Twitter research: disciplines, methods, and ethics. Aslib Journal of Information Management, 66 (3), 250-261.