# Data-Driven Park Planning†

## -Comparative Study of Survey with Social Media Data-

Jisoo Sim
Associate Research Fellow, Korea Research Institute for Human Settlements

## Ⅰ. Introduction

Big data is emerging as a potentially valuable source of data and may come to play an important role in urban interpretation and planning. For example, sensing urban lands using social media provides the characteristics of urban use, urban area, and urban imaginaries (Kelley, 2013; Lee et al., 2013; Frias-Martinez and aias-Martinez, 2014). Social media data in particular has the potential to improve our understanding of human needs and desires (Afzalan and Muller, 2014; Georgiou, 2017) and their attitudes toward a place and its attributes (Hollander & Shen, 2017). Because users post their activities and feelings on social media to communicate (Kankanhalli et al., 2005), social media has the potential to be used in design and planning as a tool for understanding the needs of the people who use the landscape. Social media data also don't have geographical limitations, which means researchers can collect data from users worldwide at their desks. Finally, social media data allows researchers to be predict future trends from past datasets.

This study focuses on these two issues: (1) the emergence of big data, especially social media data, and its ability to reveal hidden relationships between environments and behaviors in landscape architecture, and (2) the increasing redevelopment from brownfields to parks, especially linear parks. by comparing big data analytics and small data analytics.

## Ⅱ. Literature Review

Several urban studies researchers have compared traditional and social media analytics empirically. Wang, Jin, Liu, Li, and Zhang (2016) compared social media data and survey data to assess the attractiveness of a park. They extracted data from two major social media platforms in China, Dazhongdianping and Mafengwo, and collected 5,440 posts from April 2013 to October 2015. Then they conducted an online survey asking park visitors about the attractive items in the park and collected 243 responses from 10 April to 15 April, 2016(Wang, et al., 2015). To compare the attractiveness judgments from the two sources, they divided the social media posts into positive, neutral, and negative and identified keyword from each sentiment group. After that, they analyzed the survey and classified attractiveness factors into core attractions, important attractions, and marginal attractions, then compared the ordering of the attractiveness factors (Wang et al., 2016). The findings from the social media data and survey data overlapped, they found, and reflected the attractiveness of the park well. The authors also pointed out several differences: (1) A large volume of social media data does not predict the quality of the data. (2) Social media provides unstructured data. (3) Surveys are well suited to gathering valuable information. They concluded that social media and survey data complement each other by correcting each other's weaknesses (Wang et al., 2016).

## Ⅲ. Research Design

The study is divided into three phases focused on different methodologies: (1) surveys and statistical analyses, (2) social media data and text mining, and (3) comparison of these two. Through these three phases, the study provides practical guidelines for selecting data and analytics for each type of linear park.

In phase one, I investigate park visitor behaviors and attitudes through a survey. The responses are analyzed by using statistical analyses such as PCA, correlation, and ANOVA to address both linear parks' visitor characteristics

† : 이 연구는 저자의 박사학위 논문을 요약한 것이며 더 자세한 내용은 박사학위 논문을 참조할 것.

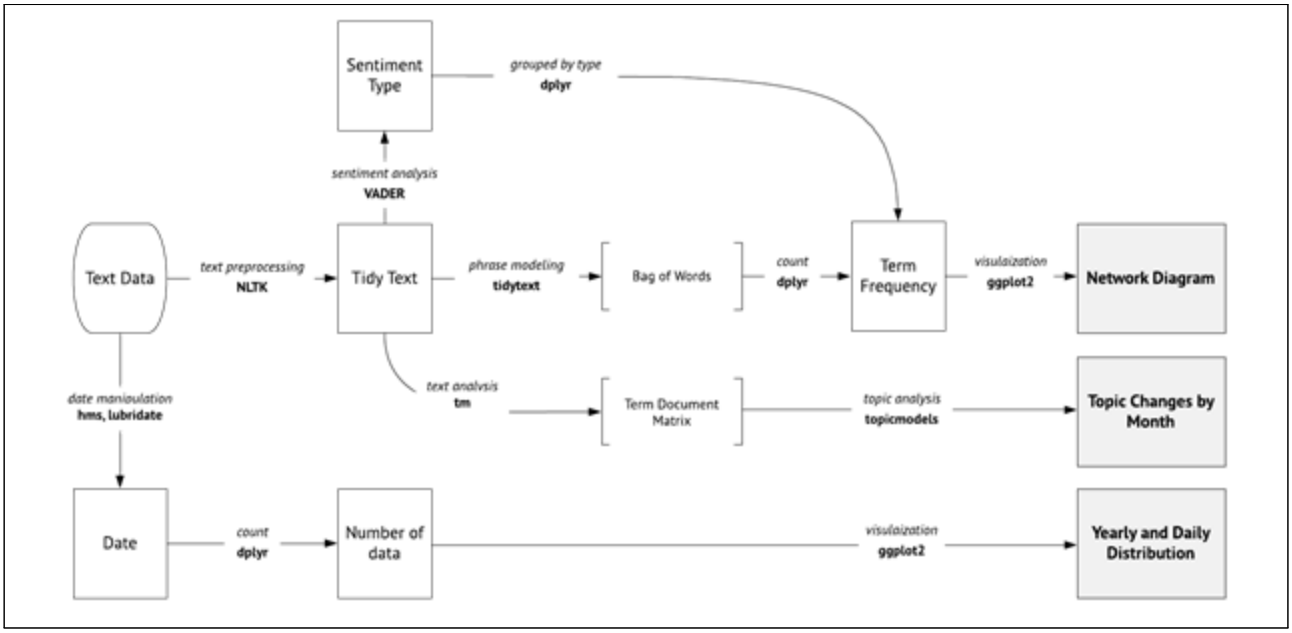


Figure 1. Research design for social media analytics

and the effectiveness of surveys. In phase two, social media analytics are used to investigate the same facts. The object is to use social media data for planning and managing linear parks. Although these differ from other parks such as neighborhood parks, amusement parks, and national parks, few studies have focused on linear parks' physical attributes instead of on their users.

Three linear parks were chosen for this study, for the following reasons. First, they have similar backgrounds, opening dates, and purposes: all three were built on unused railroads and opened between 2009 and 2015 to provide green spaces to the public. Second, all three have unique characteristics in terms of location, attributes, users, and functions. For example, one provides a recreation place, another offers a space for the arts, and the last allows visitors to be in contact with nature. By reviewing and comparing the three, I can determine what kinds of data are valuable for each type. Third, many studies have examined the economic benefits of replacing disused railroads with linear parks, but they have not investigated other aspects of these parks, such as who visits them and what they do.

The survey was conducted from April 18 to May 12, 2019, in the linear parks. Sampling dates were selected to include at least two weekdays and two weekend days. To collect more data, an extra survey was conducted for the High Bridge Trail in Farmville from May 18 to 19, 2019. I conduct three main analyses: (1) sentiment analysis, (2) topic analysis, and (3) network analysis. Sentiment analysis is intended to detect public attitudes toward the linear parks (Figure 1). Using the VADER model (Hutto & Gilbert, 2014), I assign each post a sentiment score and sentiment type: negative, neutral, or positive.

## Ⅳ. Results

In topic changes by month, the High Line is more monotonous than the 606. Figure 2 shows the results of the topic analysis. People using the High Line mentioned "art," "walk," "photo," "people," and "city" frequently. It can be presumed that one reasons for the park's success is New York City itself; art could be another reason. High Line users like taking photos in all seasons, and people mentioned "photo" frequently. This means the High Line is a place to memorialize by posting on social media, not an everyday place. The urban views on the High Line may also be loved; High Line users mentioned "building" and "tower" frequently.

This study compares two methods of data collection, focused on activities and benefits. The survey asked respondents to check all the activities they did in the park. Social media users' activities were detected by term frequency in social media data. Both results ordered the activities similarly. For example social interaction and art viewing were most popular on the High

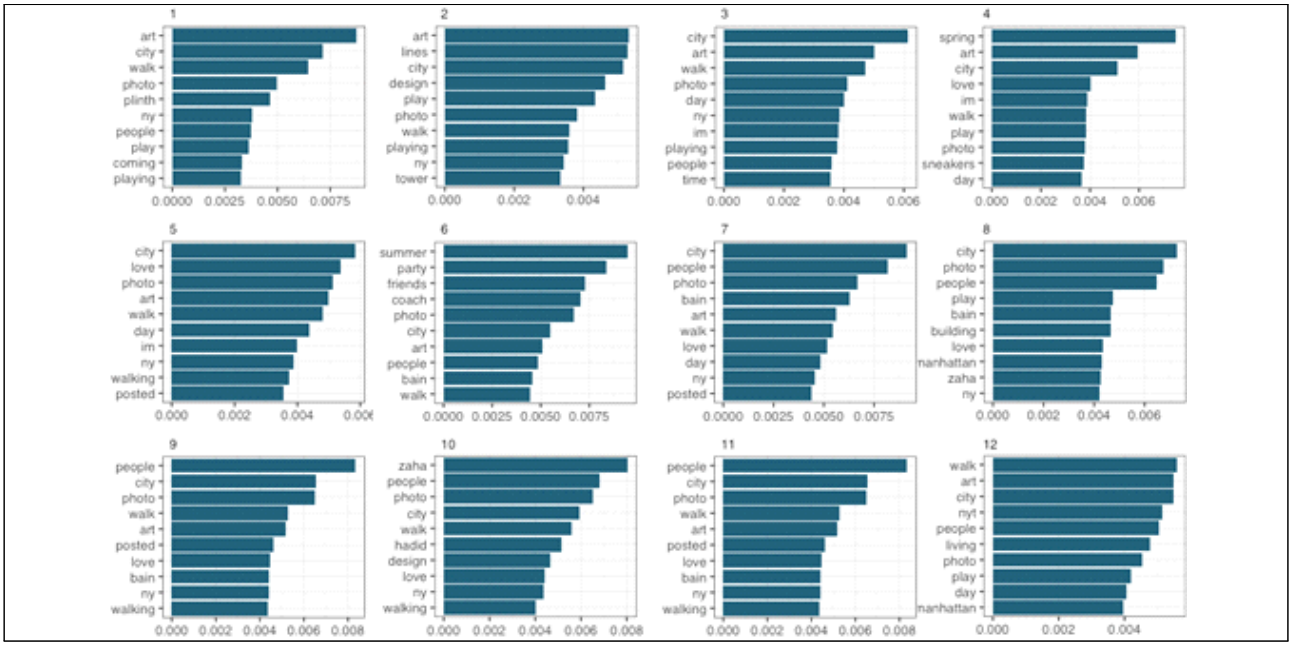


Figure 2. Topic analysis of the high line by month (x: topic/total tweets, y: topic).

Line, then the 606, then the High Bridge according to both methods. Both methods also reported that High Line visitors engaged in viewing from overlooks the most. As for benefits, according to both methods vistors to the 606 were more satisfied than High Line visitors with the parks' social and natural benefits. These results suggest social media analytics can replace surveys when the textual information is sufficient for analysis.

Social media analytics also differ from surveys in accuracy of results. For example, social media revealed that 606 users were interested in events and worried about housing prices and crimes, but the pre-designed survey could not capture those facts. Social media analytics can also catch hidden and more general information: through cluster analysis, we found possible reasons for the High Line's success in the arts and in the New York City itself. These results involve general information that would be hard to identify through a survey.

On the other hand, surveys provide specific information and can describe visitors' demographics, motivations, travel information, and specific benefits. For example, 606 users tend to be young, high-income, well educated, white, and female. These data cannot be collected through social media.

## References

1. Afzalan, N., and B. Muller(2014). The Role of social media in green infrastructure planning: a case study of neighborhood participation in park siting. Journal of Urban Technology 21(3): 67-83.
2. Frias-Martinez, V., and E. Frias-Martinez(2014). Spectral clustering for sensing urban land use using Twitter activity. Engineering Applications of Artificial Intelligence 35: 237-245.
3. Georgiou, T.(2017). Understanding the Real World through the Analysis of User Behavior and Topics in Online Social Media (Ph.D., University of California, Santa Barbara).
4. Hollander, J. B. and Y. Shen(2017). Using Social Media Data to Infer Urban Attitudes About Bicycling: An Exploratory Case Study of Washington DC. In Springer Optimization and Its Applications. City Networks pp. 79-97.
5. Hutto, C. J., and E. Gilbert(2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International AAAI Conference on Weblogs and Social Media. Presented at the Eighth International AAAI Conference on Weblogs and Social Media.
6. Kankanhalli, A., B. C. Tan and K. K. Wei(2005). Contributing knowledge to electronic knowledge repositories: an empirical investigation. MIS Quarterly: 113-143.
7. Kelley, M. J.(2013). The emergent urban imaginaries of geosocial media. GeoJournal 78(1): 181-203.
8. Lee, R., S. Wakamiya and K. Sumiya(2013). Urban area characterization based on crowd behavioral lifelogs over Twitter. Personal and Ubiquitous Computing 17(4): 605 - 620.
9. Wang, D., G. Brown, Y. Liu and I. Mateo-Babiano(2015). A comparison of perceived and geographic access to predict urban park use. Cities, 42, 85-96.
10. Wang, R., J. Zhao and Z. Liu(2016). Consensus in visual preferences: the effects of aesthetic quality and landscape types. Urban Forestry & Urban Greening, 20: 210-217.