



Article Understanding an Urban Park through Big Data

Jisoo Sim * D and Patrick Miller

Landscape Architecture Program, College of Architecture and Urban Studies, Virginia Tech, 800 Drillfield Dr., Blacksburg, VA 24060, USA; pmiller@vt.edu

* Correspondence: jisoosim@vt.edu

Received: 30 July 2019; Accepted: 3 October 2019; Published: 10 October 2019



Abstract: To meet the needs of park users, planners and designers must know what park users want to do and how they want the park to offer different activities. Big data may help planners and designers gain this knowledge. This study examines how big data collected in an urban park could be used to identify meaningful implications for planning and design. While big data have emerged as a new data source, big data have not become an accepted source of data due to a lack of understanding of big data analytics. By comparing a survey as a traditional data source with big data, this study identifies the strengths and weaknesses of using big data analytics in park planning and design. There are two research questions: (1) what activities do park users want; and (2) how satisfied are users with different activities. The Gyeongui Line Forest Park, which was built on an abandoned railway, was selected as the study site. A total of 177 responses were collected through the onsite survey, and 3703 tweets mentioning the park were collected from Twitter. Results from the survey show that ordinary activities such as walking and taking a rest in the park were the most common. These findings also support existing studies. The results from social media analytics found notable things such as positive tweets about how the railway was turned into a park, and negative tweets about diseases that may occur in the park. Therefore, a survey as traditional data and social media analytics as big data can be complementary methods for the design and planning process.

Keywords: urban park; onsite survey; user analysis; big data; social media analytics; sentiment analysis

1. Introduction

Although big data have emerged as a critical source of data and are playing an essential role in urban studies, it is still uncommon in park planning and design. Understanding urban settings by using big data can reveal heretofore hidden characteristics of urban areas [1–3]. The use of social media data, a type of big data, has the potential to provide a deeper understanding of human attitudes and perceptions toward urban places [4] and explain resulting behaviors [5,6]. There are significant strengths of social media data such as indisputability, volume, and real-time data. Since social media users post their opinions on their social media platform [7], social media data represent public opinions more directly than other traditional methodologies when researchers try to understand public attitudes and perceptions. Social media can generate enormous amounts of data. In the case of Twitter, over 473,400 tweets are posted every minute around the world, and 2.5 quintillion bytes of data are created every single day [8]. Social media data can also be collected in real-time [9]. While other researchers have been attuned to the advantages of big data in social science research, researchers in landscape architecture and urban planning rarely use big data analytics.

One study used social media data to identify successful public spaces [10], and another one used location-based data to trace the human visit dynamics in parks [11]. However, due to the difficulties in measuring social media data [12] and concern over how closely the content analysis can link to that context [13,14], different content analysis such as sentiment and frequency analysis are rarely used.

The approach used most often in the past, or the traditional approach, for designing and planning open spaces is called "the demand approach" [15]. The demand approach uses the stated desires of people from interviews or questionnaires and provides recommendations of recreation and amenities to meet those demands. As stated above, most parks, urban forests, and greenways are planned and designed using the demand approach [16–20]. If done properly, these methods require considerable time and expense to collect the needed data.

One of the reasons for the reluctance in using social media data is that the reliability of social media analytics has not been verified in the design and planning. Therefore, this study examined how social media data can be used to understand: (1) visitor activities in parks and (2) visitor satisfaction about the park. This was done by comparing two different methods: (1) a survey as traditional data and (2) social media analytics as big data. The survey will examine how people indicate they use a park, then collected tweets mentioning the park will be analyzed. After that, the results derived from the two approaches will be compared to identify the similarities and differences between the two methods.

This article contains five sections: introduction, literature review, methods, results, and conclusion. The literature review answers the benefits that park visits provide to users, how to assess user activities and their attitudes, and the pros and cons of the survey and social media data. In the methods section, we address the study site, data collection, and analytic methods used. The results section provides our findings. We then discuss the conclusions and implications in the conclusions.

2. Literature Review

This section describes how urban parks are used and compares two different methods that can be used to understand how people use urban parks. Parks benefits can be grouped into three categories: economic, health, and environmental benefits [21,22]. This section describes the social benefits of urban parks and how urban parks contribute to social interactions. Then, we discuss traditional analytics including surveys and big data analytics focused on social media analytics in urban studies. After that, we compare the pros and cons of the two methods.

2.1. Social Benefits and Social Interaction in Urban Parks

Several terms have been used in studies of the social benefits from urban parks. Many studies have investigated the role of urban parks as an essential space to increase the quality of daily life of an urbanized society [23,24]. Empirical evidence supports that urban parks, greenways, and urban forests in urban contexts improve the quality of life in many ways. In addition to environmental benefits such as pollution purification, urban parks contribute to improving health, enhancing social interactions, and providing peacefulness [25–28]. Social benefit is a crucial metric to determine the success of a redevelopment project. Social benefit is related to how the project generates urban vitality. Urban vitality can be measured by social interactions [29–32]. Social interaction, as a process of interactivity among more than two people and the relationship between people and spaces, includes all forms of communication such as cooperation, competition, playing, informing, negotiating and bargaining, and creates the placeness [30–32]. Furthermore, urban vitality can be considered in various ways. Jane Jacobs describes urban vitality as a place to provide chances for good relations between people [33]. Jalaladdini and Oktay (2012) states that urban vitality is a safer, more desirable, and more attractive place that offers more choices for social activities [34]).

Ulrich [27], for example, compares psychophysiological reactions toward three landscape types: landscapes with vegetation, with water, and with urban content. By using metrics that represent physiological and psychological measures, he found that vegetation and water landscapes had greater beneficial influences on the positive psychological feelings of park users. Chiesura [26] states that urban parks provide social and psychological benefits and enrich our lives with meaning and emotions. She explains that being in nature makes people feel positive. To do a better job of planning and designing urban parks that will offer positive emotions, we need to understand the tools or approaches that can be used. These include the role of traditional methods and big data approaches. Peters, Elands, and Buijs [35] assert that urban parks can be a trigger to generate social interactions by stimulating social cohesion. They used a survey, observations, and interviews to carry out social interaction research in five urban parks in the Netherlands. According to this study, urban parks can promote the mingling of different ethnic groups and encourage interactions among visitors. Coley, Sullivan, and Kuo [36] verified that urban parks encouraged social interactions among residents and contributed to social cohesion. Their findings revealed that natural elements increased opportunities for social interactions, and using outdoor places promoted communication within neighborhoods [36]. Natural environments can also help people relax as well as promote social interactions [37]. Kuo et al. [37] showed that the presence of greenspaces in the inner city generated positive responses from residents. The density of urban nature may also increase the sense of safety. These studies show practical evidence that urban parks not only foster social interaction, but also contribute to the sense of safety in urban environments. These studies show the practical implications that urban parks contribute to urban vitality by fostering social interactions.

Age groups have been considered as an important predictor of park visits, especially for people aged over 50 years who visit parks and participate in park activities at a lower rate than other age groups [38–40]. Numerous studies have stated that physical activities and visits to a park significantly decrease with age. Another sociodemographic characteristic is gender. In 1997, Portes [41] indicated that gender represents a major dimension of social structure and offers important insights to understand many phenomena. In more recent research, Reed, Price, Grost, and Mantinan [42] found significant differences in gender and park use. Social media data can answer questions by detecting social networks revealed on a social media platform [43]. Social network analysis contributes to detecting communities [44], social roles, and social cohesion [43].

2.2. Traditional Analytics for Understanding People

There are several methods to understand how people use a place and how people interact with each other in a place. Researchers have developed several ways to measure these variables through field observation, surveys, and interviews. Since these methods have been used in studies of urban parks for a long time, these methods are often called traditional analytics [45,46]. Some studies have termed these methods as small data analytics when compared to new methods such as big data analytics [47]. In this study, we call these methods traditional analytics to stress their historicity.

Among the traditional analytics such as an interviews, observations, and focus group meetings, surveys can be considered as the representative method for studying urban parks and behaviors [48,49]. A survey can satisfy two main concerns of traditional methods: fairness and efficiency. As many surveys are targeted to anticipated park users and often obtain data about the survey participants, it is easy to verify who the participants are and the extent to which they are representative of park users. In many studies, a survey is used to understand people's attitudes and their behavior in urban open spaces such as parks. Peters, Elands and Buijs [35] used a survey to find that the urban parks can promote social cohesion. Whiting, Larson, Green, and Kralowec [50] used a park visitors survey to identify motivation and preferences for outdoor recreation.

While traditional assessment tools can be used to evaluate the on-site benefits of parks through preset surveys, interviews, and observations [51–54], the weakness of these methods is that the intention of the researcher can be reflected in the questions asked and in the wording of the questions. One limitation of many surveys is that they are self-selecting. Participation is voluntary. Those who are motivated to participate may not represent all park users. Another limitation is that surveys require the researcher put the questionnaire together to preconceive what types of activities users may want to participate in. If certain activities are not included in the survey, they will not appear in the results. A survey can provide good data when it includes activities of interest to the participants. How a survey compares to big data analytics will be discussed next.

2.3. Big Data Analytics as New Techniques for Understanding Park Usage

Big data sources, especially social media data, hold the potential for enhancing the understanding of human behaviors and perceptions of urban places. Social media such as Facebook, Twitter, Instagram, and Flickr are widely used by people to post and share their opinions and communicate with each other [7]. Social media can be considered important source data to identify and understand how people interact within urban spaces. Although accessing social media data is not free, and researchers often have to pay fees for the data or licensing agreement, it still has considerable strengths. There are three reasons as to why social media could be a valuable data source: (1) social media users post text to express their thoughts directly [55]; (2) collecting social media data allows researchers to trace the past data, and (3) social media data are cheap data and their volume is enormous when compared to traditional data. According to social cognitive theory (SCT), it explains how people memorize in terms of three things: (1) their understanding of the activities; (2) their participation; and (3) the physical environmental [56]. Users post information on social media to share their thoughts and communicate with each other [7]. When we considered the motivation of using social media based on SCT, there are two kinds of factors: intrinsic (personal) and extrinsic (environmental) factors. In terms of the extrinsic factors, user behavior is affected by people sharing information [7]. Big data, especially social media data, may be considered as representative media to capture a user's thoughts and behaviors.

Sentiment analysis is one way to evaluate one's perceptions and emotions, whether they are positive or negative [57–60] by using social media data. Among the sentiment analysis classification, the lexicon-based approach was used in this study. The lexicon-based approach uses positive and negative sentiment terms by using a dictionary based-approach [61]. The algorithm gives one point to the positive words and subtracts one point for every negative word. Then, all scores are evaluated for total content. The algorithm gives a zero score to content that has no positive (higher than 0) or negative words (lower than 0) or offsets positive and negative words in the content. The score refers to the users' attitude toward a park. The sentiment score derived from the algorithm shows whether the users were satisfied with their experience of the park. If the score is bigger than zero, the score is counted as a positive experience. If the score is under zero, it represents a negative experience.

Although big data are accepted as reliable data, there are several challenges that need to be overcome when using big data. First, big data require a researcher to have specific computer skills to access and analyze the data. To write a code that uses software packages requires that a researcher is knowledgeable about computer program languages such as R, Python, Java, and C [62]. Second, it is common for some researchers to believe that big data are available to all without the need to pay a fee [63]. Accessing big data sources is not free. Social media data from Twitter, and Facebook require the terms of services (ToS) to be continuously updated and changed to protect privacy issues [9]. Compared to big data, traditional data sources such as census data are more freely and widely accessible and more available for everyone to analyze with standard software. Third, there is a lack of standardization of methods for collecting and analyzing big data. While some studies very precisely address data collection and methods [64], other studies are rather vague [65,66]. Fourth, big data analytics will contain sampling errors. In the case of traditional methods such as a survey or interview, there are standard methods to minimize the sampling error [67,68]. However, big data, especially social media data, have challenges beyond data collection [69]. Users who post on social media are not likely to be representative of all users. While social media may provide meaningful implications about people and their behavior in an urban setting [9,70] in terms of its volume, social media data also have limitations that it may be biased from the opinions of younger people [69]. Hargittai (2015) pointed out that big data studies have methodological challenges of limited sampling frames from those inclined to use social media platforms. For example, Facebook users are younger, and the results from that platform tend to reflect younger opinions [71].

2.4. Comparison of Traditional Methods with Big Data Analytics

Survey data as a traditional method of collecting data are strong as the sampling process can be easier when compared to big data analytics. In terms of a sample, an onsite survey may select a visitor by contacting every fifth or tenth visitor for validity in a sampling process [67,68]. However, since researchers cannot identify the demographic information of social media users, it is difficult to control sampling validity in big data analytics [69].

Compared to traditional data, social media data provide valuable information about the behavior of people in a specific space. While social media data have limitations related to a lack of generalization to the entire population [72], social media data may have other advantages. First, there is an enormous number of users around the world who publish their daily activities. Second, social media data include a variety of types of information such as trivial travel experiences that cannot be documented with traditional methods. When we considered the motivation of using social media based on social cognitive theory, there are two types of factors: intrinsic (personal) and extrinsic (environmental) factors. In terms of the extrinsic factors, the users' behaviors are affected by people sharing information [7].

3. Methodology

This section describes the study site, how data were collected, and the analytics used. The authors selected the Gyeongui Line Forest Park in Seoul, Korea, which has emerged as an urban hotspot to examine how visitors use the park through a survey and social media analysis. Since the park is surrounded by four universities, it has become a hotspot for young generations in Korea. For collecting data, a survey as a traditional method and social media as big data were selected. Among the traditional methods, a survey covers larger samples [67] and represents targeted information that is designed by researchers. Social media data also cover large samples and social media platforms are used to post and share the thoughts and daily activities of social media users [6]. Several studies have compared survey data and social media data [73,74]. This study selected a survey and social media data for comparison. At the park, an onsite survey was conducted on weekdays and weekends. Twitter postings that mentioned the park title were also collected during this period. With the survey and social media data, statistical analysis and sentiment analysis were conducted to derive results.

3.1. Study Site

As above-mentioned, the Gyeongui Line Forest Park was selected as the study site. This park was used as a railroad for the last 100 years (Figure 1a). When the railroad was turned into an underground railroad, the site was vacant and abandoned. As one of the urban revitalization projects, the Seoul Metropolitan Government built the park on the vacant land. The Seoul Metropolitan Government expected the park to facilitate the economic growth of adjacent areas [75]. From Mapo-gu to Yongsan-gu, Seoul, the park, 6.3 km in length, crosses the center of the city. The park was divided into three phases, and each phase was completed in 2012, 2015, and 2016, respectively [75].

At present, the park has become the center of younger generations because of its location. Four universities, Yonsei, Hongik, Seokang, and Ewha Women's University are located within a 700 m radius from the park (Figure 1b). The park has emerged as an urban hotspot where young people gather and frequently visit [76]. Well-known districts for young people, Sinchon, Hongdae, and Ewha districts, are also located near the park.



Figure 1. (a) Rail road system of Korea; (b) Gyeongui Line Forest Park.

3.2. Data Collection

The data were collected using two different methods: a survey and social media (Table 1). For the survey data, an onsite survey was conducted and a total of 192 samples were collected. For the social media data, postings mentioned in the keywords related to the park were collected and a total of 3703 tweets were filtered.

Methods	Date	Sample Size	Respondents
Park visitor survey	08.15–08.18. 2018 10.01–10.08. 2018	177 ¹	Park visitors (older than 18)
Social media data	06.10–09.20. 2018	3703	Keywords: 'Gyeongui Line', 'Gyeongui Line Forest Park' and 'Yeontral Park'

¹ Responses to the different methods of data collection.

3.2.1. Survey

The survey was used as the traditional method for comparison with social media data and to understand user activities in the park and user satisfaction about the park. Among the traditional methods, surveys can be considered as a way that satisfies two things: fairness and efficiency. Fairness is related to concepts such as democracy, representativeness, transparency, and public acceptability [77]. This concept concerns the perceptions of participants and the public and whether public participation has been conducted in a manner that accurately reflects the views of the target population. The concept of efficiency refers to the ease with which data can be collected [77].

The survey was conducted from August 15–18 and October 1–8, 2018. The sampling dates at the park were selected to include weekdays and weekends. At the park, the researcher approached every third adult visitor and asked of their willingness to take the park survey. A total of 192 respondents stated they were willing to participate in the survey, and 15 were disqualified because they did not complete the survey.

The questionnaire was divided into five sub-parts: park visiting frequency and activities, satisfaction of the visit, social interactions, social cohesions, and respondent information. The first

part included information about the visits to the park (frequency, duration, company) and activities visitors did in the park [78]. Nine activities in the park were carefully selected following the previous literature [24,78,79]. User satisfaction included questions about the visitors' feelings toward the park features and their experience ranged from '1 = strongly disagree' to '5 = strongly agree'. For the social interaction, participants were asked to answer several variables including feeling safe, participating in social programs in the park, and forming new relationships with others. Since social cohesion occurs based on trusting social relationships such as a sense of community [80], the social cohesion part of the survey asked about the nature of trusting relationships with those in their neighborhood. Following Perez et al. (2015), survey participants were asked to self-report about social cohesion factors among residents and their willingness to take action for the common good. Based on questions from the literature, the authors also added a question about the residents' attitudes toward and willingness to make social relationships with other visitors of the park.

3.2.2. Social Media Data

For the social media data, all tweets mentioned in the keywords related to the park were collected from July to September 2018 through Twitter API. Every tweet that mentioned one of three keywords, 'Gyeongui Line', 'Gyeongui Line Forest Park', and 'Yeontral Park' were scrawled. Three keywords were selected as the park has different names: some call it the 'Gyeongui Line Forest Park', others refer to it as 'Gyeongui Line', while the younger generation know the park as 'Yeontral Park'. All tweets in the dataset were posted inside South Korea. A total of 3703 postings were collected over three months.

3.3. Data Analytics

3.3.1. Survey Analytics

Data analytics was designed to answer two main research questions: (1) what activities in the park could be detected through the survey and (2) how many visitors were satisfied with their experiences in the park. For analysis, the authors mainly used R Studio, which is a free and open source tool and an integrated development environment (IDE) for the R language [81]. Given its accessibility and the ease in using the R programming language, it was decided in this paper to use R.

For the survey data, descriptive statistics and correlation analysis were conducted to identify how visitors used the park and how satisfied they were with their park experiences. The surveyed data collected included user activities, user satisfaction, and social interaction. A frequency analysis was also conducted. After the frequency analysis, the relationship between satisfaction and social interactions was analyzed by using correlation analysis.

3.3.2. Social Media Analytics

Regarding social media data, two main analytics were conducted to investigate the data. Text mining and sentiment analysis can be used to identify the implications. Text mining is usually divided into three areas: data mining, statistics, and linguistics, which aims to extract meaningful information from unstructured textual data [82]. The emergence of social media applications has contributed to the growth of text mining usage. We used Hu and Liu's (2004) approach to conduct sentiment analysis [83]. Sentiment analysis was then used to understand the preferences and attitudes of users.

Frequency analysis was conducted to identify major words used in social media posts. Content analysis was then conducted on the posted texts to categorize them into single word categories. These categories were then counted to identify the main reason for the users' sentiments. These sentiments can represent people's opinions, sentiments, emotions, and attitudes [84]. This analysis provided the researchers with user assessments of the park and the reasons for those sentiments.

4. Results

The results were divided into three parts: the survey, social media, and a comparison of these two methods. The survey results indicate that visitors occasionally visited the park for relaxation. User satisfaction showed a positive relationship with social interaction. From the survey results, the park as an urban hotspot provides a place for relaxation and fosters social interaction and cohesion. The social media results revealed the positive attitudes of park visitors toward the park, especially attributes of the park such as 'forest' and 'railroads'. The results of the social media analysis also pointed out that park visitors positively reacted to interactive activities with others. For the comparative test of the difference between the survey and social media, there were no differences between the satisfaction from the survey and positive postings from social media.

4.1. Survey Results

The survey data revealed the characteristics of the park visitors. Like in previous studies, the park tended to be visited by more females (54.24%) and young people under 40 years of age (64.41%). The majority of visitors came from other districts and around thirty percent of visitors were residents of adjacent areas to the park (Table 2).

		.
Demographics	п	Percentage
Gender		
Male	81	45.8%
Female	96	54.2
Age		
18–29	64	36.2
30–39	50	28.3
40-49	33	18.6
50-59	21	11.9
Over 60	9	5.1
Residents ¹		
Yes	54	30.5
No	123	69.5

Table 2. Characteristics of visitors.

¹ Residents who live within 800 m of the park.

The descriptive statistics of the survey explained the basic uses of the park. Table 3 provides an understanding of the visits. Many visitors came to the park less than once a month (29.9%) or were visiting the park for the first time (23.7%). Additionally, some portion of visitors visited the park for more than two days (14.1%) a week or every day (10.7%). This result supports that park visitors use the park occasionally rather than on a daily basis, and that visitors normally came to the park on special occasions rather than regular visits.

Numbers of Visit	Percentage
Every day	10.7%
More than 2 days a week	14.1
Once a week	9.6
1–3 times a month	11.9
Less than once a month	29.9
This is the first time	23.7

Table 3. Frequency of visit.

In terms of the stated desire for visiting and actual use of the park, Table 4 shows the results. Activities in four categories—physical activities, mental health, social interaction, and

other activities—were asked and the results are stated in Table 4. Visitors came to the park to refresh their daily life (53.7%) and take a rest (36.2%). The main activity in this park differs slightly from existing studies that have identified physical activities as the main activities in a park [85]. The highest percentage of visitors used the park to interact with their friends or family (50.9%). Under 20 percent of visitors used this park to do physical activity. The average satisfaction score was based on a scale of 1 = rarely satisfied to 5 = highly satisfied. Users of the park were slightly satisfied (mean = 3.66), but the level of satisfaction was not that high. For social interaction in the park, users answered that the park contributed to reinforcing social interaction (mean = 3.56). The social cohesion satisfaction of adjacent community residents was 3.51. The satisfaction score was compared between two groups: visiting alone and visiting the park with friends or family. Pearson's Chi-squared test was used to examine the relationship between factors where there was no difference between the two groups (*p*-value = 0.49). One of the results of the survey was the correlation analysis. The correlation between satisfaction and social interaction was positive (correlation coefficient = 0.61, *p*-value < 0.0001). Although many researchers have verified that social interaction is a social benefit from a park [32,35], the relationship between the park visitors' satisfaction and social interaction level has rarely been identified.

Activity	Stated Desire	Actual Use
Physical activities		
Biking	0.0%	0.0%
Walking or running	18.1	19.2
Mental health		
Refresh one's mind	53.7	27.1
Relax and restoration	36.2	31.1
Social interaction		
Seeing others	26.6	41.7
Having time with friends or family	17.0	50.9
Other activities		
Just passing through the park	36.7	72.3
Enjoying hobby (photo, sports, etc.)	3.4	5.1

Table 4. Motivation and the actual activities in the park.

Next, the correlation analysis was conducted once again to identify the relationship between satisfaction (sati), social interaction (soc), and social cohesion (soco) in the park (Figure 2). Figure 2 represents the correlation between these factors. The graphs with the red line refer to the correlation graphs between satisfaction (sati) and social interaction (soc), social interaction (soc) and social cohesion (soco), and satisfaction (sati) and social cohesion (soco). The numbers represent the correlation coefficient. The relationships between the factors of satisfaction, social interaction, and social cohesion were positive. For example, satisfaction and social interaction had a positive relationship (correlation coefficient = 0.62). There was also a positive relationship between satisfaction and social cohesion (0.32), but the relationship that was not strong. Social interaction and social cohesion relationship also showed a positive relationship that was not strong (0.42). This means that visitors of the park who were satisfied with their experience tended to value social benefits such as increasing social interaction or social cohesion.



Figure 2. Correlation analysis between factors. Abbreviations: sati, satisfaction; soc, social interaction; soco, social cohesion.

4.2. Social Media Results

As mentioned in Section 2, the sentiment analysis of social media data provides insight into the users' attitude toward the park. The results indicate that more users had a positive experience than those who had a negative one (Figure 3). Figure 3 shows the number of tweets and sentiment scores of each tweet during the study period. The results show that social media users generally reacted positively when they posted about the park. This also supports the previous study that social media users tend to highlight their positive attributes [86].



Figure 3. The result of the sentiment analysis (from negative sentiment (red, less than 0) to positive sentiment (blue, higher than 0)).

To figure out the type of activities that made the positive or negative sentiment, the dataset divided into two sub-sets: the positive data, and the negative data. Then, the word frequency analysis was comprised to extract the main reasons for each sentiment. The most frequently mentioned words in the positive data were related to food and eating such as 'beer', 'cake', 'coffee', 'cooked', 'lamb', 'tteokbokki (street-food in Korea)', and 'taste'. Physical features of the park were also mentioned in the dataset: 'forest', and 'railroad'. This indicates that the unique character of the park is important to the positive experience of users. Words which implied the relationship between people were also mentioned frequently such as 'people', 'we', and 'you'. (Figure 4a). From these results of the most frequently mentioned words, we can conclude that among the park visitors, social media users tended to have a positive experience about the unique attributes of the park such as 'forest' and 'railroad' and also tended to enjoy casual treats in the park. Furthermore, 'railroad' and 'railway' refer to the previous railway on the grounded Gyeonggui Line.



Figure 4. (a) Word frequency analysis of positive sentiments; (b) Word frequency analysis of negative sentiments.

The negative sentiments were more complex in real-life than the positive sentiments, since the result of the frequency analysis indicated that the number of users who mentioned a word was too low to draw any conclusions. In other words, each word was mentioned 20 times. Only some words such as 'park', 'it', 'I', and 'people' were mentioned over 20 times in the negative dataset. Other words that were mentioned between 10 to 20 times were 'disease', 'tsutsugamushi (disease from lawn)', and 'ghastly'. All three words were related to the disease from the lawn. This data show that park visitors worried about the disease when they visited the park (Figure 4b).

In terms of activities in the park, social media data indicate that eating something with someone most often resulted in positive sentiment toward the park. This assumption was derived from the most frequently mentioned words such as 'eating', 'foods', and 'we'. One of the more interesting results is the use of personal pronouns. When the users represented a negative sentiment, they used 'I' rather than 'we'. From the use of pronouns, it seems that experiences with others tended to be positive while negative experiences tended to seen as relating to the individual. These insights from social media analytics could not be found in the survey results. Since the survey questionnaire was designed to identify participation in traditional activities in the park, the questionnaire did not capture the importance of park activities such as 'eating'. However, social media data tended not to capture ordinary activities in the park such as 'walking' and 'relaxing'. Perhaps this is because people take something like walking for granted. When people posted about their activities and experiences on Twitter, they did not upload ordinary, common activities of their daily life and preferred instead to share their more extraordinary or less common events with others.

4.3. Comparing the Survey and Big Data

Results from both the social media data and the survey data indicate a mostly positive experience in the park. Even though there were some differences between users of the park and the respondents of the survey, their experiences in the park tended to be positive.

The results from the social media data and the survey data explained the users' activities, satisfaction (sentiments), and their social interaction in the park. The survey data and the social media data represent the different aspects of the users' activities. For example, the social media data could catch the extraordinary events of the daily lives of users such as eating delicious food with someone. The survey data included the ordinary, common activities in the park such as walking, relaxing, and chatting. In terms of social interaction, both methodologies support the idea that the park can contribute to social interaction. In terms of the social interactions, social media data and the survey data suggest that a park is a place to reinforce social interactions. According to the survey, users of the park indicated that the park contributed to social interaction and social cohesion. This result regarding the social interaction and satisfaction was identified as a positive relationship, however, the causal factors between two variables were not verified.

5. Conclusions

From the survey and social media, this study attempted to figure out the uses and meaning of urban parks. To overcome the problem that the survey may not reach sufficient users, social media can be used to fill a gap between a preset questionnaire and the reality. Through two lenses, the authors found meaningful implications. Through the survey, the results showed that park visitors mainly used the park for restoration and relaxation from their daily life. Additionally, park visitors who were more satisfied with their park experience also experienced more social cohesion and enhanced social interaction. Through the social media data, the authors found that social media users who visited and posted about the park enjoyed unique attributes or features of the park such as the railway, railroad, and forest features. Furthermore, social media users reacted positively to relatively small and common experiences such as coffee, beer, and casual foods in the park and negatively to less common and potentially threatening experiences such as a potential health and disease issues.

This study provides meaningful insights into the possible use of social media content as the data source for landscape architecture research. According to the study, the social media data and big data analytics can be used to detect a new activity type in a park. For the social interaction, park visitors mentioned 'we' when they posted positive contents instead of 'I'. This may mean that social media users tend to share their status when they are with others.

This study also identified several limitations to the use of social media data in park studies, especially in terms of sampling challenges. By choosing a study site where young generations tend to visit, the authors tried to minimize a gap between social media users and real visitors. It may diffuse the challenges, but is not a perfect solution. The sampling challenges of social media data have to carefully consider when researchers use social media data for future research.

This study is a starting point in identifying more advanced research methods that can be used to augment existing methods. This paper makes several significant contributions to landscape architecture studies by providing a way to use social media data as a tool for understanding neighborhoods and resident preferences. Through the comparison of these methodologies, this study identified the pros and cons of the methods and explored the possibility of using big data analytics for understanding urban parks and the people who use them.

Author Contributions: Study Design, Writing—original draft, J.S.; Study Design, Writing—review & editing, P.M. Funding: This research received no external funding.

Acknowledgments: We are grateful to the anonymous referees for sharing their valuable and helpful comments. Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Frias-Martinez, V.; Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. *Eng. Appl. Artif. Intell.* **2014**, *35*, 237–245. [CrossRef]
- 2. Kelley, M.J. The emergent urban imaginaries of geosocial media. *GeoJournal* 2013, 78, 181–203. [CrossRef]
- 3. Lee, R.; Wakamiya, S.; Sumiya, K. Urban area characterization based on crowd behavioral lifelogs over Twitter. *Pers. Ubiquitous Comput.* **2013**, *17*, 605–620. [CrossRef]
- 4. Hollander, J.B.; Shen, Y. Using Social Media Data to Infer Urban Attitudes About Bicycling: An Exploratory Case Study of Washington DC. In *City Networks*; Springer Optimization and Its Applications: Cham, The Netherlands, 2017; pp. 79–97. ISBN 978-3-319-65336-5.
- 5. Afzalan, N.; Muller, B. The Role of Social Media in Green Infrastructure Planning: A Case Study of Neighborhood Participation in Park Siting. *J. Urban Technol.* **2014**, *21*, 67–83. [CrossRef]
- 6. Georgiou, T. Understanding the Real World through the Analysis of User Behavior and Topics in Online Social Media. Ph.D. Thesis, University of California, Santa Barbara, CA, USA, 2017.
- 7. Kankanhalli, A.; Tan, B.C.; Wei, K.-K. Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Q.* **2005**, *29*, 113–143. [CrossRef]
- 8. Ahmad, I. How Much Data Is Generated Every Minute? [Infographic]. Available online: https://www. socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/ (accessed on 15 September 2018).
- 9. Poorthuis, A.; Zook, M. Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. *J. Urban Technol.* **2017**, *24*, 115–135. [CrossRef]
- 10. Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities* **2017**, *64*, 66–78. [CrossRef]
- Hamstead, Z.A.; Fisher, D.; Ilieva, R.T.; Wood, S.A.; McPhearson, T.; Kremer, P. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Comput. Environ. Urban Syst.* 2018, 72, 38–50. [CrossRef]
- 12. Hausmann, A. Creating 'buzz': Opportunities and limitations of social media for arts institutions and their viral marketing. *Int. J. Nonprofit Volunt. Sect. Mark.* **2012**, *17*, 173–182. [CrossRef]
- Gonçalves, P.; Araújo, M.; Benevenuto, F.; Cha, M. Comparing and Combining Sentiment Analysis Methods. In Proceedings of the First ACM Conference on Online Social Networks, Boston, MA, USA, 7–8 October 2013; pp. 27–38.
- 14. Riffe, D.; Lacy, S.; Fico, F.; Watson, B.; Lacy, S.; Fico, F.; Watson, B. *Analyzing Media Messages: Using Quantitative Content Analysis in Research;* Routledge: Abingdon, UK, 2019; ISBN 978-0-429-46428-7.
- Cordell, H.K.; Betz, C.; Bowker, J.M.; English, D.B.K.; Mou, S.H.; Bergstrom, J.C.; Teasley, R.J.; Tarrant, M.A.; Loomis, J. Outdoor Recreation in American life: A National Assessment of Demand and Supply Trends. In *Champaign, IL: Sagamore Publishing, xii*; Sagamore Publishing: Champaign, IL, USA, 1999; p. 449.
- 16. French, J.S. *Urban Green: City Parks of the Western World;* Kendall/Hunt Publishing Company: Dubuque, IA, USA, 1973; Volume 619.
- 17. Heckscher, A.; Robinson, P.C. *Open Spaces: The Life of American Cities*; HarperCollins Publishers: New York, NY, USA, 1977.
- 18. Maruani, T.; Amit-Cohen, I. Open space planning models: A review of approaches and methods. *Landsc. Urban Plan.* **2007**, *81*, 1–13. [CrossRef]
- 19. Turner, T. Landscape planning: A linguistic and historical analysis of the term's use. *Landsc. Plan.* **1983**, *9*, 179–192. [CrossRef]
- 20. Turner, T. Open space planning in London: From standards per 1000 to green strategy. *Town Plan. Rev.* **1992**, 63, 365. [CrossRef]
- 21. Hoshino, T.; Kuriyama, K. Measuring the Benefits of Neighbourhood Park Amenities: Application and Comparison of Spatial Hedonic Approaches. *Environ. Resour. Econ.* **2010**, *45*, 429–444. [CrossRef]
- 22. De Ridder, K.; Adamec, V.; Bañuelos, A.; Bruse, M.; Bürger, M.; Damsgaard, O.; Dufek, J.; Hirsch, J.; Lefebre, F.; Pérez-Lacorzana, J.M.; et al. An integrated methodology to assess the benefits of urban green space. *Sci. Total Environ.* **2004**, 334, 489–497. [CrossRef] [PubMed]

- 23. McCormack, G.R.; Rock, M.; Toohey, A.M.; Hignell, D. Characteristics of urban parks associated with park use and physical activity: A review of qualitative research. *Heal. Place* **2010**, *16*, 712–726. [CrossRef] [PubMed]
- 24. Campbell, L.K.; Svendsen, E.S.; Sonti, N.F.; Johnson, M.L. A social assessment of urban parkland: Analyzing park use and meaning to inform management and resilience planning. *Environ. Sci. Policy* **2016**, *62*, 34–44. [CrossRef]
- 25. Barbosa, O.; Tratalos, J.A.; Armsworth, P.R.; Davies, R.G.; Fuller, R.A.; Johnson, P.; Gaston, K.J. Who benefits from access to green space? A case study from Sheffield, UK. *Landsc. Urban Plan.* **2007**, *83*, 187–195. [CrossRef]
- 26. Chiesura, A. The role of urban parks for the sustainable city. Landsc. Urban Plan. 2004, 68, 129–138. [CrossRef]
- 27. Ulrich, R.S. Natural Versus Urban Scenes: "Some Psychophysiological Effects". *Environ. Behav.* **1981**, *13*, 523. [CrossRef]
- 28. Ulrich, R.S.; Simons, R.F.; Losito, B.D.; Fiorito, E.; Miles, M.A.; Zelson, M. Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* **1991**, *11*, 201–230. [CrossRef]
- 29. Cranshaw, J.; Schwartz, R.; Hong, J.; Sadeh, N. The livehoods project: Utilizing social media to understand the dynamics of a city. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–7 June 2012.
- Hari, R.; Kujala, M.V. Brain basis of human social interaction: From concepts to brain imaging. *Physiol. Rev.* 2009, *89*, 453–479. [CrossRef] [PubMed]
- 31. Moulay, A.; Ujang, N. Legibility of Neighborhood Parks and Its Impact on Social Interaction in a Planned Residential Area. *Archnet-IJAR: Int. J. Arch. Res.* **2016**, *10*, 184–194. [CrossRef]
- 32. Moulay, A.; Ujang, N.; Said, I. Legibility of neighborhood parks as a predicator for enhanced social interaction towards social sustainability. *Cities* **2017**, *61*, 58–64. [CrossRef]
- 33. Jacobs, J. The Death and Life of Great American Cities; Vintage Books: New York, NY, USA, 1961.
- 34. Jalaladdini, S.; Oktay, D. Urban Public Spaces and Vitality: A Socio-Spatial Analysis in the Streets of Cypriot Towns. *Procedia-Soc. Behav. Sci.* **2012**, *35*, 664–674. [CrossRef]
- 35. Peters, K.; Elands, B.; Buijs, A. Social interactions in urban parks: Stimulating social cohesion? *Urban For. Urban Green.* **2010**, *9*, 93–100. [CrossRef]
- 36. Coley, R.L.; Sullivan, W.C.; Kuo, F.E. Where Does Community Grow? The Social Context Created by Nature in Urban Public Housing. *Environ. Behav.* **1997**, *29*, 468–494. [CrossRef]
- 37. Kuo, F.E.; Bacaicoa, M.; Sullivan, W.C. Transforming Inner-City Landscapes: Trees, Sense of Safety, and Preference. *Environ. Behav.* **1998**, *30*, 28–59. [CrossRef]
- 38. Fisher, K.J.; Li, F.; Michael, Y.; Cleveland, M. Neighborhood-Level Influences on Physical Activity among Older Adults: A Multilevel Analysis. *J. Aging Phys. Act.* **2004**, *12*, 45–63. [CrossRef]
- Li, F.; Fisher, K.J.; Brownson, R.C.; Bosworth, M. Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *J. Epidemiol. Community Heal.* 2005, 59, 558–564. [CrossRef]
- 40. Payne, L.; Orsega-Smith, B.; Godbey, G.; Roy, M. Local parks and the health of older adults. *Parks Recreat.* (*Ashburn*) **1998**, *33*, 64–70.
- 41. Portes, A. Immigration Theory for a New Century: Some Problems and Opportunities. *Int. Migr. Rev.* **1997**, *31*, 799–825. [CrossRef] [PubMed]
- 42. Reed, J.A.; Price, A.E.; Grost, L.; Mantinan, K. Demographic Characteristics and Physical Activity Behaviors in Sixteen Michigan Parks. *J. Community Health* **2012**, *37*, 507–512. [CrossRef] [PubMed]
- 43. Hansen, D.; Shneiderman, B.; Smith, M.A. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*; Morgan Kaufmann: Burlington, MA, USA, 2010; ISBN 978-0-12-382230-7.
- 44. Gibbons, J.; Nara, A.; Appleyard, B. Exploring the imprint of social media networks on neighborhood community through the lens of gentrification. *Environ. Plan. B Urban Anal. City Sci.* **2018**, 45, 470–488. [CrossRef]
- 45. Aoki, Y. Evaluation Methods for Landscapes with Greenery. Landsc. Res. 1991, 16, 2–6. [CrossRef]
- 46. Eiter, S.; Vik, M.L. Public participation in landscape planning: Effective methods for implementing the European Landscape Convention in Norway. *Land Use Policy* **2015**, *44*, 44–53. [CrossRef]

- Carr, S.J.; Lassiter, A. Big Data, Small Apps: Premises and Products of the Civic Hackathon. In Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics; Springer International Publishing: Cham, Switzerland, 2017; ISBN 978-3-319-40902-3.
- 48. Buchinger, S.; Kriglstein, S.; Brandt, S.; Hlavacs, H. A survey on user studies and technical aspects of mobile multimedia applications. *Entertain. Comput.* **2011**, *2*, 175–190. [CrossRef]
- 49. Yarrow, C. A Preliminary Survey of the Public's Concept of Amenity in British Forestry. *Forestry* **1966**, *39*, 59–67. [CrossRef]
- 50. Whiting, J.W.; Larson, L.R.; Green, G.T.; Kralowec, C. Outdoor recreation motivation and site preferences across diverse racial/ethnic groups: A case study of Georgia state parks. *J. Outdoor Recreat. Tour.* **2017**, *18*, 10–21. [CrossRef]
- 51. Ewing, R. Using a visual preference survey in transit design. *Public Works Manag. Policy* **2001**, *5*, 270–280. [CrossRef]
- 52. Birch, E.L. The Observation Man, A Study of William, H. Whyte 1986. Available online: https://repository. upenn.edu/cgi/viewcontent.cgi?article=1021&context=cplan_papers (accessed on 10 October 2019).
- 53. Carmona, M. *Public Places, Urban Spaces: The Dimensions of Urban Design;* Architectural Press: Boston, MA, USA, 2010.
- 54. More, T.A.; Stevens, T.; Allen, P.G. Valuation of urban parks. Landsc. Urban Plan. 1988, 15, 139–152. [CrossRef]
- 55. Edelman, R.; Hirshberg, P. Public RelationSHIPS: Communications in the Age of Personal Media. Available online: https://www.scribd.com/document/6576091/Blogger-Survey-FINAL-FromprinterRVSD (accessed on 10 October 2019).
- 56. Bandura, A. Social cognitive theory of self-regulation. *Organ. Behav. Hum. Decis. Process.* **1991**, *50*, 248–287. [CrossRef]
- Barbosa, L.; Feng, J. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China, 23–27 August 2010; pp. 36–44.
- 58. Go, A.; Huang, L.; Bhayani, R. Twitter sentiment analysis. Entropy 2009, 17, 252.
- 59. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation, Valletta, Malta, 17–23 May 2010.
- 60. Schweitzer, L. Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter. *J. Am. Plan. Assoc.* **2014**, *80*, 218–238. [CrossRef]
- Maynard, D.; Funk, A. Automatic Detection of Political Opinions in Tweets. In *Semantic Web: ESWC 2011* Workshops; García-Castro, R., Fensel, D., Antoniou, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 88–99.
- 62. Jenkins, T. On the difficulty of learning to program. In Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences, Citeseer, Loughborough, UK, 27–29 August 2002; pp. 53–58. Available online: http://www.psy.gla.ac.uk/~{}steve/localed/jenkins.html (accessed on 10 October 2019).
- 63. Arribas-Bel, D. Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Appl. Geogr.* **2014**, *49*, 45–53. [CrossRef]
- 64. Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77. [CrossRef]
- 65. Kay, S.; Zhao, B.; Sui, D. Can Social Media Clear the Air? A Case Study of the Air Pollution Problem in Chinese Cities. *Prof. Geogr.* 2015, *67*, 351–363. [CrossRef]
- 66. Widener, M.J.; Li, W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Appl. Geogr.* **2014**, *54*, 189–197. [CrossRef]
- 67. Kalton, G.; Graham, K. Introduction to Survey Sampling; SAGE: Newcastle upon Tyne, UK, 1983; ISBN 978-0-8039-2126-9.
- Larson, R.; Csikszentmihalyi, M. The Experience Sampling Method. In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*; Csikszentmihalyi, M., Ed.; Springer The Netherlands: Dordrecht, The Netherlands, 2014; pp. 21–34. ISBN 978-94-017-9088-8.
- 69. Choudhury, M.D.; Lin, Y.-R.; Sundaram, H.; Candan, K.S.; Xie, L.; Kelliher, A. How Does the Data Sampling Strategy Impact the Discovery of Information Diffusion in Social Media? In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23–26 May 2010.

- Agrawal, D.; Budak, C.; El Abbadi, A.; Georgiou, T.; Yan, X. Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks. In *Databases in Networked Information Systems*; Madaan, A., Kikuchi, S., Bhalla, S., Eds.; Springer International Publishing: New York, NY, USA, 2014; pp. 1–16.
- 71. Hargittai, E. Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *ANNALS Am. Acad. Polit. Soc. Sci.* 2015, 659, 63–76. [CrossRef]
- 72. Chou, W.-Y.S.; Hunt, Y.M.; Beckjord, E.B.; Moser, R.P.; Hesse, B.W. Social media use in the United States: Implications for health communication. *J. Med. Internet Res.* **2009**, *11*, e48. [CrossRef]
- 73. Brindley, P.; Cameron, R.W.; Ersoy, E.; Jorgensen, A.; Maheswaran, R. Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health. *Urban For. Urban Green.* **2019**, *39*, 45–54. [CrossRef]
- 74. Wang, Z.; Jin, Y.; Liu, Y.; Li, D.; Zhang, B. Comparing Social Media Data and Survey Data in Assessing the Attractiveness of Beijing Olympic Forest Park. *Sustainability* **2018**, *10*, 382. [CrossRef]
- 75. Seoul Metropolitan Government Seoul Solution. Available online: https://seoulsolution.kr/ko/node (accessed on 15 September 2018).
- 76. Oh, H.-J.; Lee, M.-H. The difference between residents and visitors on the influencing factors of satisfaction in urban linear park: A case study of the Gyeongui-line forest park's Yeonnam-dong section. *J. Korea Plan. Assoc.* 2018, 53, 207. [CrossRef]
- 77. Rowe, G.; Frewer, L.J. A Typology of Public Engagement Mechanisms. *Sci. Technol. Hum. Values* **2005**, *30*, 251–290. [CrossRef]
- Kaźmierczak, A. The contribution of local parks to neighbourhood social ties. *Landsc. Urban Plan.* 2013, 109, 31–44. [CrossRef]
- 79. Gehl, J.; Svarre, B. How to Study Public Life; Island Press: Washington, DC, USA, 2013.
- 80. Perez, L.G.; Arredondo, E.M.; McKenzie, T.L.; Holguin, M.; Elder, J.P.; Ayala, G.X. Neighborhood social cohesion and depressive symptoms among Latinos: Does use of community resources for physical activity matter? *J. Phys. Act. Heal.* **2015**, *12*, 1361–1368. [CrossRef] [PubMed]
- 81. Schumacker, R.E. *Learning Statistics Using*, R.; SAGE Publications: Thousand Oaks, CA, USA, 2014; ISBN 978-1-4833-2477-7.
- 82. He, W.; Zha, S.; Li, L. Social media competitive analysis and text mining: A case study in the pizza industry. *Int. J. Inf. Manag.* **2013**, *33*, 464–472. [CrossRef]
- Hu, M.; Liu, B. Mining and Summarizing Customer Reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; ACM: New York, NY, USA, 2004; pp. 168–177.
- 84. Liu, B. Sentiment Analysis, Opinion Mining, Emotion Classification; Cambridge University Press: Cambridge, UK, 2015.
- 85. Godbey, G.; Mowen, A. The Benefits of Physical Activity Provided by Park and Recreation Services: The Scientific Evidence. *Australas. Park. Leis.* **2011**, *14*, 26–29.
- 86. Ellison, N.; Heino, R.; Gibbs, J. Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *J. Comput. Mediat Commun.* **2006**, *11*, 415–441. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).