Supplementary Information

1. Data collection and pretreatment

Three data sets with georeferenced microblogs or reviews of urban park records and comments from the most popular tourist attraction SMD platforms with multi-types in China (Liang and Zhang, 2021), i.e., DZDP (http://www.dianping.com), Ctrip (http://you.ctrip.com), and Weibo (https://weibo.com), were composed to study the patterns of urban park visit sentiment in Shanghai. All data used in this paper are publicly available. The content of urban park reviews and microblogs of the study sites, including the park name, check-in or time post and site, user ID and name, and text from the travel record and comment, were collected by application programming interfaces within the study periods from 2018.5.1 to 2019.4.30. The raw review and microblog texts for 300 parks obtained from three SMD platforms were filtered and cleaned up by deleting blank and duplicated records and georeferenced wrong data. A dictionary was built and used for text treatment. It comprised of HowNet (http://www.keenage.com/)—one of the most famous and popular online common-sense knowledge based on Chinese words (Ma et al., 2020; Su et al., 2021)—as the basic lexicon, and the vocabulary of Internet catchphrases and terms associated with travel and Shanghai city from the lexicon of Sogou Pinyin Input (https://pinyin.sogou.com/dict/)—one of the most famous and popular Chinese input methods (Ren and Hong, 2017)—as the supplementary lexicon. It was implemented in Python version 3.6 programming language (https://www.python.org/) for part-of-speech tagging, word segmentation, and stop word removal with a Chinese stop-work list including stop words, such as punctuation, numbers, and location names.

2. Sentiment analysis

Considering microblogs/reviews on such SMD platforms were short texts, lexicon-based approach, a mature and reliable method for Chinese sentiment analysis and especially good at short texts, was applied in this study (Chen et al., 2019; Su et al., 2021). With the sentiment dictionary from HowNet mentioned above, the words in each review or microblog after pretreatment were compared to the emotional lexicon and assigned scores according to the degree of emotion expressed by the words in the dictionary. The 21 scores ranging from −10 to 10 were assigned to each word. The scores above, equal to, and below zero represent positive, neutral, and negative sentiment, respectively. The higher the absolute of the positive or negative score, the stronger the positive or negative sentiments indicated by the words. The sentiment score for each review or microblog (SSRM) was the total score for all the emotion words in the text of the review or microblog. Considering that the visitor repeating their reviews or microblogs might amplify their sentiment, the reviews/microblogs of the same park posted by the same visitor on the same day (RMPVD) are merged into one piece by calculating their average sentiment score (SSRMPVD). For park *p*, visitor *v*, and day *d*, $\mathrm{SSRMPVD}_{pvd}$ was calculated using Equation (1):

$$\mathrm{SSRMPVD}_{pvd} = \mathrm{Sum}(\mathrm{SSRM}_{pvd}) \,/\, \mathrm{Num}(\mathrm{RM}_{pvd}) \qquad (1)$$

where $\mathrm{Num}(\mathrm{RM}_{pvd})$ is the number of reviews/microblogs that visitor *v* posted on day *d* for park *p*.

3. Data sorting and screening

The scores of RMPVD with SSRMPVD from −100 to 100 are regarded as more reasonable and rational, and remained for later statistical analysis. After sorting and screening, a total of 363,036 pieces of RMPVD within the study period for Shanghai parks were obtained. Among them, 68,541, 217,722, and 76,773 pieces were from Ctrip, DZDP, and Weibo, respectively. These pieces of RMPVD from three SMD platforms were applied for descriptive statistics of the overall distribution of sentiment scores for all the parks in Shanghai. Then, the urban park sites with an RMPVD count of under 50, which might reduce sentiment accuracy calculated later for the urban park due to limited sample data, were removed. The RMPVD of parks covering all three platforms remained to analyze park sentiment patterns using different SMD sources. Finally, 201,004, 64,461, and 64,128 pieces of RMPVD from the platforms DZDP, Ctrip, and Weibo, respectively, covering 115 urban parks, were obtained.

4. Statistical analysis of sentiment patterns

Pieces of RMPVD from the three platforms obtained above were used to calculate the average sentiment score by park (SSPVD) and average park sentiment scores on each day (SSDVD). For park *p* and day d, $SSPVD_p$ and $SSDVD_d$ were calculated using Equations (2) and (3):

$$SSPVD_p = Sum(SSRMPVD_p) / Num(SSRMPVD_p) \quad (2)$$

$$SSDVD_d = Sum(SSRMPVD_d) / Num(SSRMPVD_d) \quad (3)$$

where $Num(SSRMPVD_p)$ and $Num(SSRMPVD_d)$ are the piece numbers of RMPVD for park p and on day d, respectively. The temporal distribution of park sentiment was studied through descriptive statistics using different time periods, such as year, workday/non-workday, and season. Day proportions of $SSDVD_d$ above, equal to, and below 0 were also calculated and used to analyze the distributions of daily positive, neutral, and negative emotions for different time periods.

With SSPVD obtained, ArcGIS 10.7 (ESRI, Redlands, CA, USA) was used to study Shanghai's park sentiment spatial distribution from the three SMD sets. It provided point maps for the geographic distribution of SSPVD to present high and low areas. The Moran's I index was used to measure whether the SSPVDs for Shanghai parks were spatially autocorrelated through Global Moran's I. The Getis-Ord Gi* statistic used the General G index to measure the degree of clustering for either the high or low values to identify spatial distribution patterns with statistically significant hot and cold spots of SSPVD in the study area. To explore whether the distribution of park sentiment has a significant correlation with the time period, the Spearman's coefficients were applied to analyze the bivariate correlation between variables 'season time unit' and 'park average sentiment score' in each season (SSPSVD), and between variables 'workday/non-workday time unit' and 'park average sentiment score on a workday or non-workday' (SSPWVD). Differences of SSPSVD for each season from the season average (DSSPSVD) and differences of SSPSVD between a workday and non-workday (DSSPWVD) were calculated and used to study how the spatial characteristics vary over time using point maps. For park p, season s, workday w, and non-workday nw, $SSPSVD_{ps}$, $SSPWVD_{pw}$, $SSPWVD_{pnw}$, $DSSPSVD_{ps}$, and $DSSPWVD_{pw}$ were calculated using Equations (4)–(8):

$$SSPSVD_{ps} = Sum(SSRMPVD_{ps}) / Num(SSRMPVD_{ps}) \quad (4)$$

$$SSPWVD_{pw} = Sum(SSRMPVD_{pw}) / Num(SSRMPVD_{pw}) \quad (5)$$

$$SSPWVD_{pnw} = Sum(SSRMPVD_{pnw}) / Num(SSRMPVD_{pnw}) \quad (6)$$

$$\text{DSSPSVD}_{ps} = \text{SSPSVD}_{ps} - \text{Sum}(\text{SSPSVD}_{p}) / 4 \qquad (7)$$

$$\text{DSSPWVD}_{pw} = \text{SSPWVD}_{pw} - \text{SSPWVD}_{pnw} \qquad (8)$$

where Num(SSPWVD$_{pw}$), Num(SSPWVD$_{pw}$), and Num(SSPWVD$_{pnw}$) are the piece number of RMPVD for park p in season s, on workday w, and non-workday nw, respectively. Sum(SSPSVD$_{p}$) is the sum of the SSPSVD of park p for all four seasons.

5. Statistical analysis of related factors and SMD comparisons

A series of potential influential factors (Table S1), including internal factors indicating the attributes of the park itself and external factors indicating the aspects existing around, were selected based on previous studies (Fan et al., 2021; Guo et al., 2019; Liu and Xiao, 2021; Lyu and Zhang, 2019; Zhang and Zhou, 2018) to explore their impacts on visitors' sentiment in Shanghai parks for the three SMD sets. For the independent variables of the factors and the response variable of the SSPVD, multiple linear regressions (MLRs) were used to investigate the relationships with visitors' sentiment in parks and examine the level of sentiment influence of each factor. The variance inflation factor (VIF) was calculated for each variable to test the multicollinearity among the variables, as some of them were correlated. Since there were enough sample numbers for the parks involved, variables of factors with VIF less than 5 were multicollinear, ignorable, and remained for continued MLR. To compare each potential factor's influence on the response variable, all explanatory variables were normalized before the regression. The standardized coefficients (beta coefficients) determined the relative importance of the factors – the larger the absolute value, the more important the factor. Moreover, Spearman's correlation coefficients were used to measure the relationships between each pair of the three SMD sources for the variables of SSPVD, SSPSVD, and SSPWVD at different time periods. These statistical analyses were conducted in SPSS Version 26 (IBM, Armonk, NY, USA).

Table S1 Factors that might be related to park sentiment.

Type	Name	Description
Internal	Size	Land area of the park.
	Scenic spot count	The number of scenic spot point of interests (POIs) in the park.
	Star rating	The star rating with five levels (i.e., 5, 4, 3, 2, and 1 star) published every year by the government presenting the quality of the park (http://lhsr.sh.gov.cn/). These are based on various factors for rational and comprehensive considerations of park quality, such as park area, facilities, and scenery (Liang et al., 2017; Liang and Zhang, 2018).
	Attraction grade	The attraction grade with five levels (i.e., 5, 4, 3, 2 and 1 A) published in the notice announcement issued in the official site by the Ministry of Culture and Tourism of the People's Republic of China (https://www.mct.gov.cn/) based on various factors for rational and comprehensive considerations of tourist attraction quality, such as attraction landscape, and service and management.
	Service facilities	The number of POIs of service facilities, such as restrooms, restaurants, and a tourist service center in the park.
	Visit number	The number of SMD visits to the park.
	Visit density	The density of SMD visits to the park.
	Entrance fee	The entrance fee of the park.
	Online reputation	The reputation scores or grades on SMD platforms for the park.
External	Distance from the urban center	The distance from an urban park to the center of the city (Fan et al., 2021).
	Traffic convenience	The number of traffic facilities, such as bus and subway stations, parking lots, and access around the urban park within 800 m (Brown et al., 2014; Guo et al., 2019; Liang and Zhang, 2018).
	Commercial facilities	The number of commercial facilities, such as shops, hotels, and restaurants around an urban park.
	Residential places	The number of residential communities around the park.
	Employment places	The number of employment places around the park.

References cited

Brown, G., Schebella, M.F., Weber, D., 2014. Using participatory GIS to measure physical activity and urban park benefits. Landscape and Urban Planning 121, 34-44.

Chen, J., Becken, S., Stantic, B., 2019. Lexicon based Chinese language sentiment analysis method. Computer Science and Information Systems 16, 639-655.

Fan, Z., Duan, J., Lu, Y., Zou, W., Lan, W., 2021. A Geographical Detector Study on Factors Influencing Urban Park Use in Nanjing, China. Urban Forestry & Urban Greening, 126996.

Guo, S., Yang, G., Pei, T., Ma, T., Song, C., Shu, H., Du, Y., Zhou, C., 2019. Analysis of factors affecting urban park service area in Beijing: Perspectives from multi-source geographic data. Landscape and Urban Planning 181, 103-117.

Liang, H., Chen, D., Zhang, Q., 2017. Walking accessibility of urban parks in a compact megacity. Proceedings of the Institution of Civil Engineers - Urban Design and Planning 170, 59-71.

Liang, H., Zhang, Q., 2018. Assessing the public transport service to urban parks on the basis of spatial accessibility for citizens in the compact megacity of Shanghai, China. Urban Studies 55, 1983-1999.

Liang, H., Zhang, Q., 2021. Temporal and spatial assessment of urban park visits from multiple social media data sets: A case study of Shanghai, China. Journal of Cleaner Production 297, 126682.

Liu, R., Xiao, J., 2021. Factors Affecting Users' Satisfaction with Urban Parks through Online Comments Data: Evidence from Shenzhen, China. International Journal of Environmental Research and Public Health 18, 253.

Lyu, F., Zhang, L., 2019. Using multi-source big data to understand the factors affecting urban park use in Wuhan. Urban Forestry and Urban Greening 43, 126367.

Ma, Y., Ling, C., Wu, J., 2020. Exploring the Spatial Distribution Characteristics of Emotions of Weibo Users in Wuhan Waterfront Based on Gender Differences Using Social Media Texts. International Journal of Geo-Information 9, 465.

Ren, G., Hong, T., 2017. Investigating Online Destination Images Using a Topic-Based Sentiment Analysis Approach. Sustainability 9, 18.

Su, S., He, S., Sun, C., Zhang, H., Hu, L., Kang, M., 2021. Do landscape amenities impact private housing rental prices? A hierarchical hedonic modeling approach based on semantic and sentimental analysis of online housing advertisements across five Chinese megacities. Urban Forestry & Urban Greening 58, 126968.

Zhang, S., Zhou, W., 2018. Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. Landscape and Urban Planning 180, 27-35.