Supplementary Materials

Digital Expressions of Sense of Place at Cultural Heritage Sites in Ningbo on Social Media

Zhuocheng Ying, Xusheng Cheng, Dongjie Situ, Tian Ruan and Chao Gao

Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo, Zhejiang, China

Correspondence: Chao Gao (gaoqinchao1@163.com)

Scope of this file

This file provides supplementary information on data acquisition and preprocessing, PSR interaction-weight sensitivity, empirical Bayes prior-strength sensitivity, illustrative heritage sites, and the spatial distribution of digital expression types. Term frequencies, complete sensitivity outputs, platform-level indicator data, and the site-level spatial-factor and OPGD results underlying Figs. 7 and 8 are supplied as separate data files.

Data acquisition and preprocessing details

Publicly accessible posts and first-level comments were retrieved through separate keyword searches using the name of each of the 78 heritage sites on Douyin and Weibo. The archived crawler configuration used QR-code login, keyword-search mode and CSV output. Searches used each platform's default comprehensive search mode, and the crawler retrieved results in the order returned by the platform, without additional sorting by publication time or engagement. The maximum-post-count parameter was set to 800 per keyword, with a limit of 20 first-level comments per post; second-level comments were not collected. Because the result order was determined by the platforms, the retrieved corpus should not be treated as a random sample of all posts about each site.

Source files were mapped to platform, heritage site and content type, after which fields were standardized. Text normalization removed user mentions, URLs and bracketed emoji labels and normalized whitespace. Records were excluded when text was missing, empty, numeric only, shorter than two characters, longer than 300 characters, promotional or institutional/news content, low-information interaction text, or unrelated to the study's broad heritage and visitor-experience vocabulary. Duplicate records were identified using heritage site, source, normalized content and user identifier. IP-location labels were standardized, and only sites represented on both platforms were retained. User identifiers were used only during deduplication and were removed from the analytical dataset. This workflow reduced 297,171 raw records to 100,083 valid records.

Table S1 Composition of the final analytical corpus by platform and content type

Platform	Posts	Comments	Total	Share (%)
Douyin	13,157	42,388	55,545	55.50
Weibo	22,166	22,372	44,538	44.50
Total	35,323	64,760	100,083	100.00

Note. Posts include platform-level published content; comments are first-level comments. Percentages are calculated from the final corpus of 100,083 records.

Table S2 Data-cleaning and eligibility rules applied before analysis

Stage	Operation or rule	Purpose
Source mapping	Extract platform, heritage site and post/comment status from source files.	Create comparable records.
Field standardization	Map content, interaction counts, IP location and creation time to common fields.	Harmonize the two platforms.
Text normalization	Remove @mentions, URLs and bracketed emoji labels; normalize whitespace.	Reduce non-semantic noise.
Basic validity	Exclude missing, empty, numeric-only and <2-character texts.	Remove unusable records.
Content eligibility	Exclude >300-character, promotional, livestream/lottery, institutional/news and low-information texts.	Limit automated or non-analytical content.
Relevance screen	Retain records matching the broad heritage or visitor-experience vocabulary.	Maintain topical relevance.
Deduplication	Deduplicate by heritage site, source, normalized content and user identifier.	Avoid repeated observations.
Cross-platform check	Retain heritage sites represented on both Douyin and Weibo.	Support paired platform comparison.
Privacy processing	Remove user identifiers before analysis; retain only standardized coarse IP-location labels.	Reduce identifiability.

Note. The relevance vocabulary was used only as a broad corpus-eligibility screen. The visual, historical, functional and emotion dictionaries used to calculate the study variables are described in the main manuscript; complete term frequencies are provided in Supplementary Data 1.

Sensitivity of PSR to interaction weights

PSR was calculated separately for each platform before the platform-level results were compared. The primary interaction score was defined as $\ln(1 + L + 2C + 5R)$ for Douyin and $\ln(1 + L + 5C + 3R)$ for Weibo, where L, C and R denote likes, comments and reposts. The platform-specific weights were specified a priori as operational assumptions reflecting the stronger role of reposting in short-video diffusion and of commenting in discussion-oriented interaction; they were not treated as estimated or unbiased parameters. PSR rankings were then recalculated under equal, comment-priority, repost-priority, like-priority and deliberately extreme interaction-weight schemes. A sentiment-only reference removed interaction amplification and therefore represents a different construct rather than a simple reweighting.

Table S3 Interaction-weight schemes used in the PSR sensitivity analysis

Scheme	Douyin L/C/R	Weibo L/C/R	Role
Primary scheme	1 / 2 / 5	1 / 5 / 3	Platform-specific
Equal weights	1 / 1 / 1	1 / 1 / 1	Conventional alternative
Comment priority	1 / 5 / 3	1 / 5 / 3	Conventional alternative
Repost priority	1 / 2 / 5	1 / 2 / 5	Conventional alternative
Like priority	5 / 1 / 1	5 / 1 / 1	Extended test
Extreme comment priority	1 / 10 / 1	1 / 10 / 1	Extended test
Extreme repost priority	1 / 1 / 10	1 / 1 / 10	Extended test
Sentiment only	Not used	Not used	Reference without interaction amplification

Note. L/C/R denotes the weights assigned to likes, comments and reposts, respectively. The primary weights were applied separately to each platform; the alternative schemes were used to test whether the findings depended on these operational choices.

Table S4 Stability of site-level PSR rankings under alternative interaction weights

Comparison with primary scheme	Spearman ρ	Top-10 overlap (%)
Equal weights	0.9955	90
Comment priority	0.9992	100
Repost priority	0.9988	100
Like priority	0.9890	90
Extreme comment priority	0.9947	100
Extreme repost priority	0.9956	100
Sentiment only	0.7602	80

Note. All comparisons used 78 heritage sites. Under the interaction-weighted alternatives, pooled Spearman ρ values ranged from 0.9890 to 0.9992 and top-10 overlap ranged from 90% to 100%.

Table S5 Platform-specific stability of PSR rankings under alternative interaction weights

Comparison	Douyin ρ	Douyin top-10 (%)	Weibo ρ	Weibo top-10 (%)
Equal weights	0.9972	100	0.9889	100
Comment priority	0.9990	100	1.0000	100
Repost priority	1.0000	100	0.9965	90
Like priority	0.9905	90	0.9607	100
Extreme comment priority	0.9949	90	0.9949	100
Extreme repost priority	0.9986	100	0.9869	90
Sentiment only	0.7446	50	0.8942	60

Note. All comparisons used the same 78 heritage sites. Across all seven interaction-weighted specifications, the mean and median PSR were higher for Douyin than for Weibo; the mean difference ranged from 0.0880 to 0.1672 and the median difference ranged from 0.0646 to 0.1279. The sentiment-only reference is reported as a construct contrast and is not included in this directional robustness range.

Validation interpretation. The pooled and platform-specific results show that the main PSR ranking patterns were not driven by a single allocation of interaction weights. The sentiment-only reference produced lower correlations because removing interaction amplification changes the construct. These checks establish robustness to the tested specifications; they do not identify the primary weights as uniquely optimal or unbiased. Complete outputs are provided in Supplementary Data 2.

Sensitivity to empirical Bayes prior strength

The prior-strength parameters for GBI (κ) and PSR (τ) were varied across 1, 5, 10 and 20, with 10 as the primary setting. GBI rankings remained nearly unchanged, all 78 heritage sites retained the same four-type classification, and PSR rankings were effectively invariant across the tested range.

Table S6 Sensitivity of GBI, PSR and heritage-site classification to empirical Bayes prior strength

Outcome	Parameter comparison	Spearman ρ	Type consistency (%)
GBI (κ)	10 vs 1	0.9985	100
GBI (κ)	10 vs 5	0.9992	100
GBI (κ)	10 vs 20	0.9990	100
PSR (τ)	10 vs 1	0.999823	Not applicable
PSR (τ)	10 vs 5	0.999874	Not applicable
PSR (τ)	10 vs 20	0.999848	Not applicable

Note. All comparisons used 78 heritage sites. Type consistency applies to the GBI-based four-type classification; PSR does not enter that classification rule.

Validation interpretation. Relative to the primary setting of 10, changing κ or τ to 1, 5 or 20 yielded site-level Spearman rank correlations of 0.9985–0.9992 for GBI and 0.999823–0.999874 for PSR. All 78 sites retained their classifications when the percentile thresholds were recalculated for each κ setting. These results indicate stable site-level rankings and type assignments over the tested parameter range. Complete outputs are provided in Supplementary Data 3.

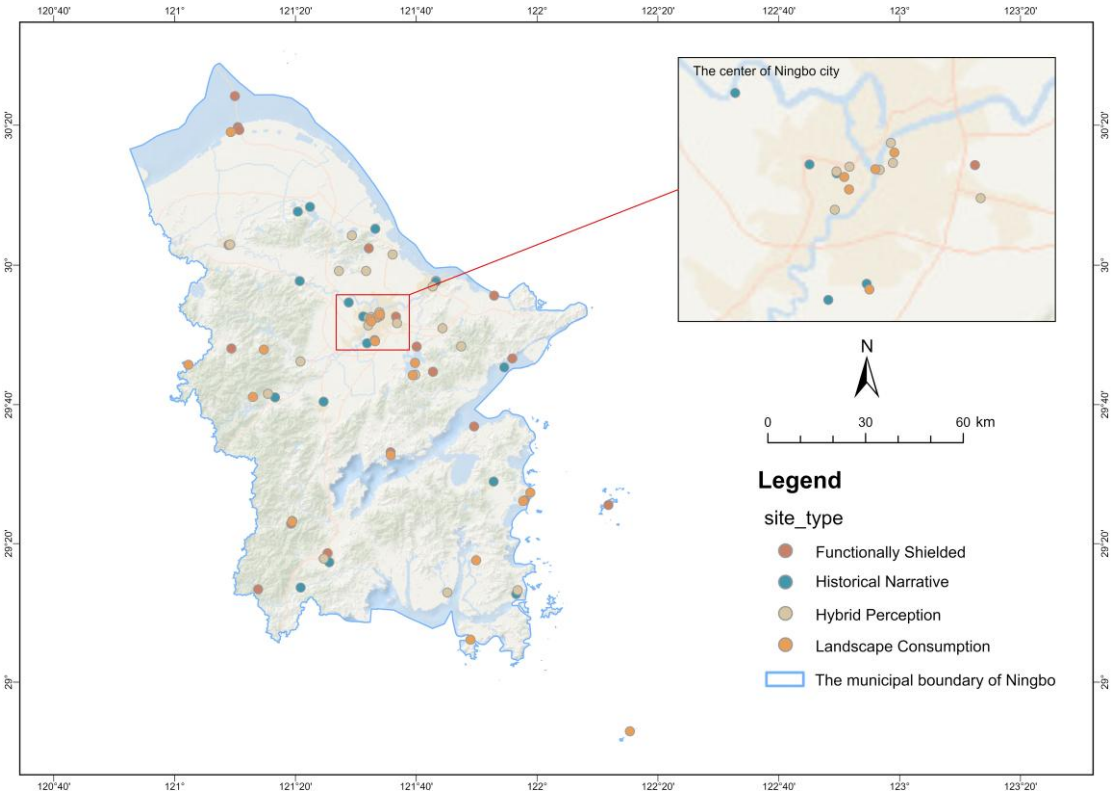
Illustrative heritage sites for the four digital expression types

Table S7 Illustrative heritage sites and classification indicators for the four digital expression types

Digital expression type	Heritage site	GBI_{site}	FSI_{site}	Main semantic characteristics
Function-shielding type	Ninghai National Mountaineering and Fitness Trail	0.631	0.678	Transportation, reservations, services and travel-guide information are prominent.
Function-shielding type	Ningbo International Conference and Exhibition Center	−0.017	0.578	Transportation, event arrangements and venue-service information are prominent.
Landscape consumption type	Xiangshan Golden Coastline	0.762	0.365	Landscape presentation, aesthetic evaluation, photography and check-in experiences are prominent.
Landscape consumption type	Yinzhou Park	0.724	0.176	Landscape appreciation and leisure experiences are prominent.
Historical narrative type	Hemudu Site	−0.802	0.050	Historical knowledge, heritage interpretation and cultural memories are prominent.
Historical narrative type	Yongfeng Granary Site	−0.766	0.071	Historical contexts and heritage interpretation are prominent.
Hybrid perception type	Gulou Cultural and Leisure Pedestrian Street	0.014	0.254	Visual expressions and historical narratives occur in relatively balanced proportions.
Hybrid perception type	Ten-Mile Red Dowry Cultural Park	−0.080	0.115	Both landscape presentation and cultural narratives are represented.

Note. GBI and FSI denote the heritage-site-level composite Gaze Bias Index and Function Shielding Index, respectively. Two sites per type illustrate high FSI, positive GBI, negative GBI or GBI close to zero. Values are rounded to three decimal places.

Spatial distribution of digital expression typologies



Supplementary Fig. S1 Spatial distribution and composition of digital expression typologies.

Spatial-factor and OPGD source data for Figs. 7 and 8

Source Data 3 provides the 78-site factor matrix used for OPGD, together with the factor-detector statistics, interaction matrices and pairwise results, optimal strata, and discretization-search outputs for FSI and GBI. Factor-detector significance is based on the optimized p values in the source results. All 15 interaction pairs for each outcome were flagged as low reliability because at least one combined stratum contained fewer than five sites; the pairwise q values are therefore presented as exploratory spatial associations.

Supplementary data file descriptions

Supplementary Data 1: Complete term-level frequencies for the visual, historical, functional and fine-grained emotion dictionaries. English glosses aid reading; Unicode-escaped original tokens preserve the exact Chinese expressions used in matching.

Supplementary Data 2: Complete PSR interaction-weight sensitivity outputs, including pooled and platform-specific rankings, top-10 overlap and cross-platform differences.

Supplementary Data 3: Complete empirical Bayes prior-strength sensitivity outputs for GBI and PSR.

Source Data 1: Indicators and summary statistics for 156 heritage-site–platform units.

Source Data 2: Paired platform-indicator matrix for 78 heritage sites.

Source Data 3: The 78-site spatial-factor matrix and the factor-detector, interaction-detector, optimal-strata and parameter-search outputs for FSI and GBI underlying Figs. 7 and 8. The interaction results are exploratory because all pairs are flagged low reliability in the source output owing to sparse interaction strata.