

Article

Digital Information Cascades and Sustainable Visitor Flow Management: Evidence from GPS Trajectories and Social Media During an Urban Festival

Yundi Wang and Zhibin Xing * 

School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China

* Correspondence: bingo@mail.swufe.edu.cn

Abstract

Urban festivals attract substantial numbers of tourists, which consequently imposes significant strain on host cities through spatial overcrowding, uneven pressure on infrastructure, and diminished quality of the visitor experience. Destination management organizations (DMOs) require effective tools to redistribute tourist flows; however, the influence of social media on tourists' actual destination choices remains insufficiently understood. We ask whether social media discussion intensity ("buzz") causally influences tourists' destination choices and whether the effect grows stronger during festivals when information asymmetry is at its peak. Combining 95,692 taxi GPS trajectories with 5995 geotagged Twitter records from the 2019 Songkran Festival in Bangkok, we constructed an exponentially weighted moving average (EWMA) buzz variable with a temporal lag that establishes causal ordering. A conditional logit model shows that district-level buzz significantly raises destination choice probability and that the effect is amplified during the festival. Causal identification rests on a triangulated strategy that combines temporal lag, placebo permutation, and Bartik shift-share instrumental variables. The festival-period IV-corrected estimate ($\hat{\beta}^{IV} = +0.019, p < 0.001$) is 51% larger than the within-period OLS estimate ($\hat{\beta}^{OLS} = +0.012, p < 0.001$), a gap consistent with classical measurement-error attenuation in sparse social-media data, and a panel 2SLS analysis at the district-day level isolates a causal visitation channel confirming that cascades reinforce spatial concentration at the tourist-flow level. The aggregate Gini coefficient of spatial concentration declines over the study window in a statistically significant monotonic trend. The positive district-level correlation between buzz and congestion does not survive district and date fixed effects, which indicates that it reflects underlying differences in attractiveness across districts rather than a direct within-district channel. These findings provide an empirical foundation for information-based visitor flow management by identifying the underlying behavioral mechanism rather than evaluating a designed intervention.



Academic Editors: Amare Nega
Wondirad and Fahimeh Hateftabar

Received: 22 March 2026

Revised: 8 May 2026

Accepted: 11 May 2026

Published: 14 May 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: information cascade; information-based visitor management; sustainable tourism; overtourism; GPS trajectory; social media

1. Introduction

Urban festivals stimulate local economic development and cultural exchange, but they also place acute sustainability pressure on host cities [1,2]. Tourist flows concentrate during festivals, degrading the environment, straining infrastructure, and reducing residents' quality of life. This constellation of pressures is increasingly framed as "overtourism" [3,4].

Bangkok's Songkran Festival exemplifies this tension. The event draws millions of domestic and international visitors each year into a handful of urban districts, where transportation gridlock and overcrowding become the rule [1]. As Milano et al. [5] argue, managing spatial imbalances is now among the central challenges of sustainable urban tourism governance.

Destination management organizations (DMOs) have traditionally relied on supply-side interventions, such as infrastructure expansion, capacity limits, and pricing mechanisms, to manage visitor flows [3,5]. A growing body of work argues that information-based instruments can redirect visitor behavior at lower political and economic cost than regulatory or pricing tools [3,6]. In today's "twin transition" era, where environmental sustainability and digital acceleration converge [7,8], social media shapes tourists' spatial decisions at an unprecedented scale [9,10]. If online conversations about particular urban areas influence where tourists actually go, then understanding the underlying behavioral mechanism is a prerequisite for any information-based visitor flow management strategy [11,12].

Three gaps remain in this fast-growing literature. First, most evidence is aggregate, examining social media's influence at the city or country level, and operational destination management needs individual-level evidence that the literature does not yet supply [13]. Second, cross-sectional and correlational designs dominate, leaving the direction of causality unresolved: tourists may post about places they have already visited rather than visit places they have seen discussed online [14,15]. Third, and most fundamentally, the interaction between information effects and situational context is theoretically underdeveloped and empirically untested in the literature. Festival periods, characterized by the disruption of familiar routines and alteration of decision heuristics, are hypothesized to be the conditions under which information cascades exert the greatest influence. However, this hypothesis has yet to be empirically tested using revealed preference data [10,16].

We address these gaps using two complementary data sources from the 2019 Songkran Festival in Bangkok: 95,692 tourist taxi GPS trajectories that record spatial choices, and 5995 geotagged Twitter posts that proxy the information environment. A conditional logit model examines how district-level social media "buzz" shapes individual destination selection among competing alternatives. Causal ordering is established through an exponentially weighted moving average (EWMA) buzz variable built on past-day activity only. The 2019 setting is also methodologically advantageous. It captures a "peak overtourism" window that serves as an unconfounded baseline of global mobility just before COVID-19. As the post-pandemic sector rebounds to historical peaks [17,18], studying cascade-driven information mechanisms under maximum-stress conditions yields robust and transferable insights for contemporary sustainable destination management.

This study makes four contributions. Theoretically, it carries information cascade theory [19] into spatial tourism, providing micro-level causal evidence that buzz raises destination choice probability and that the effect is strongest when festival-driven uncertainty is high. Empirically, it documents a spatial-scale duality in which cascades causally pull tourists toward high-buzz districts even as aggregate Gini concentration declines over the study window [3]. Methodologically, it introduces the Bartik shift-share instrumental variable strategy [20,21] to the tourism eWOM literature, along with Rotemberg and Borusyak–Hull diagnostics, establishing causal identification beyond temporal ordering and showing that measurement error in sparse social-media data substantially attenuates conventional OLS estimates. For policy, the results provide an empirical foundation for information-based visitor flow management, with the largest behavioral traction expected in search-good categories and during festival-uncertainty windows [11,12].

The remainder of this paper is organized as follows: Section 2 reviews the theoretical foundations. Section 3 describes the data and methodology. Section 4 presents the results.

Section 5 discusses the implications for sustainable tourism management, and Section 6 concludes the paper.

2. Theoretical Background

2.1. Social Media and Electronic Word-of-Mouth in Tourist Spatial Behavior

Over the past decade, the influence of social media on tourist behavior has been the subject of extensive scholarly investigation. Early work showed that user-generated content (UGC) on platforms such as TripAdvisor, Twitter, and Instagram influences destination image, trip planning, and post-visit evaluations [9,22]. Attention has since shifted from whether social media affects tourism to how and through which mechanisms digital information shapes spatial behavior on the ground [10,23]. Song and Wondirad [24] provided empirical evidence that the use of social media contributes to the spatial concentration of tourist flows within a single city. This study links online word-of-mouth to the phenomenon of overtourism at the intra-destination level.

The electronic word-of-mouth (eWOM) literature provides the primary lens for these processes. Filieri et al. [25] showed that online review credibility and volume jointly influence accommodation choice, and Liu et al. [26] found that social media engagement intensity predicts destination-selection probability. This study has two limitations constrain this literature. Most research predominantly addresses binary outcomes or satisfaction ratings rather than examining the specific locations tourists visit within a destination [15]. Among the limited studies that examine intra-destination mobility, there is a tendency to rely on descriptive spatial analysis rather than structural behavioral models [27,28]. Self-reported survey data also dominate, leaving the field vulnerable to hypothetical bias because revealed-preference evidence from GPS trajectories that capture actual spatial choices remains scarce [13].

A parallel stream uses big data analytics to study tourist mobility patterns. Over more than twenty-five years of tourist tracking research, GPS-based methods have matured from exploratory mapping tools into rigorous behavioral instruments [29]. This study shows that tourists exhibit systematic spatial preferences driven by distance, POI attractiveness, and temporal constraints [29,30]. Geotagged social media data have been used to map flows and identify popular areas [14]. Recent reviews catalogue the growing diversity of tourist-mobility tracking technologies [31], yet integrating social media and GPS data within a structural econometric framework that predicts individual spatial choice remains a methodological gap that this study addresses.

2.2. Information Cascades and Herd Behavior Under Uncertainty

Information cascade theory, first formalized by Bikhchandani et al. [19] and Banerjee [16], posits that individuals facing uncertainty rationally infer information from the actions of predecessors, sometimes overriding their own private signals. When the weight of the observed behavior exceeds that of private information, a cascade forms, and subsequent individuals follow the crowd regardless of what they know. This mechanism has been documented in financial markets, technology adoption, and consumer behavior [32].

In tourism, social media transforms cascades by making predecessors' check-ins, photos, and reviews visible at an unprecedented scale and speed [10,22]. However, this visibility is not uniformly positive. Huang et al. [10] showed that social media exposure can induce travel anxiety, so cascades carry both motivating and inhibiting potential depending on content valence and viewer disposition. When a district generates substantial online discussion, tourists may read this as a quality or safety signal and be drawn in even when the district's intrinsic attributes would not independently warrant attention. Yang et al. [23] found that exposure to peers' travel posts shifts individual venue selection toward the des-

tinations featured in those posts, and the broader literature on platform-mediated visibility suggests that online popularity metrics generate self-reinforcing concentration patterns in urban settings [33]. Crucially, the classical cascade mechanism is quantity-driven rather than valence-driven: a cascade forms when the cumulative weight of observed actions exceeds private signals, regardless of whether those actions carry positive or negative sentiment [16,19]. In the spatial-choice setting, a high tweet volume about a district signals collective attention and activity presence, which tourists may interpret as a vibrancy or safety-in-numbers cue, even when individual posts contain warnings about crowding. This attention-as-signal mechanism is conceptually distinct from the content-specific persuasion channel that sentiment analysis would capture.

Recent empirical work has confirmed the relevance of cascade theory in online tourism consumption, documenting conforming behavior among consumers who follow observed booking patterns, even when private quality signals suggest alternatives [34]. Cascade dynamics are expected to be the strongest in festival contexts. During events such as Songkran, road closures, shifted business hours, and unpredictable crowd dynamics raise information asymmetry [1,2], and under such conditions, the marginal value of external information rises, while tourists shift from deliberative to heuristic decision-making, amplifying social influence on spatial choices [16,32]. This prediction has not been tested using revealed preference data from a real festival.

2.3. Information-Based Management for Sustainable Visitor Flows

Information-based management of tourist spatial behavior draws on three conceptually distinct mechanisms that the tourism literature often conflates: informational effects (belief updating from observed signals [25]), social influence (behavior change from observation of others' actions [16,19]), and designed digital interventions (deliberate choice-architecture manipulations [6,11,35]). This study documents the first two mechanisms through organically generated buzz and does not evaluate a designed intervention.

Tussyadiah [12] positioned digital-platform-mediated interventions as a promising strategy for sustainable tourism, arguing that technological platforms can encourage pro-environmental behavior without the political costs of regulation. Recent operational work has shown that platform-level information instruments shift tourist behavior, with environmental labels on booking platforms [36], default options steering toward lower-carbon transport [37], and new-media content guiding sustainable choice [38] all demonstrating measurable effects. These instruments are designed interventions and are therefore distinct from the organically generated cascade environment analyzed here. In the spatial dimension, recent studies have proposed that real-time information provision could redistribute tourist flows within cities [39,40], but empirical evidence linking organic social media buzz to individual-level spatial choices through an identified causal pathway remains scarce.

The overtourism literature provides sustainability motivation for information-based spatial management. The spatial concentration of tourists generates negative externalities, including environmental degradation, resident displacement, and infrastructure strain [4]. Several studies have measured spatial inequality using indices such as the Gini coefficient and HHI applied to tourist distribution data [5]. More granular approaches have emerged, including GPS-based intensity-density indices that capture overtourism severity at the sub-district resolution [41] and spatiotemporal clustering algorithms that detect hotspots from mobility data [40]. Effective overtourism management requires tools that can influence tourists' spatial decisions at the individual level [3,39]. Social media information cascades represent these mechanisms. To date, no empirical study has demonstrated a causal relationship between *organically generated* social media buzz and tourists' spatial choices within a city, nor has any research evaluated whether such an influence facilitates or

impedes sustainable spatial distributions. Critically, if information cascades are organically concentrated in already-popular areas, as cascade theory predicts [19,33], then understanding the causal buzz channel is a prerequisite for designing information-based governance instruments. Any intervention that seeks to exploit or counteract the cascade mechanism must first establish the magnitude, direction, and boundary conditions of the underlying behavioral response.

2.4. Hypothesis Development

Drawing on the theoretical foundations reviewed above, we formulated four hypotheses linking social media information cascades to spatial tourist choice and sustainability outcomes.

The information cascade theory predicts that visible signals of collective behavior increase the probability of conforming actions [19]. In the spatial choice context, district-level social media buzz represents an observable signal of collective interests. Following Yang et al. [23] and the broader evidence on social media-driven visitation [33], we expect that higher buzz increases destination choice probability:

H1. *Higher lagged social media buzz about a district increases the probability that tourists choose destinations in that district.*

Cascade theory further predicts that information dependence intensifies when private information is noisy [16]. The festival periods represent exogenous uncertainty shocks, and road closures, crowd dynamics, and modified schedules diminish tourists' capacity to depend on their prior knowledge [1]. Following this logic [16,34], we hypothesize the following:

H2. *The buzz effect on destination choice is stronger during the festival period than before or after.*

The economics of information distinguishes between experience goods, whose quality can only be assessed through consumption, and search goods, whose quality can be evaluated before purchase [42,43]. Paradoxically, while experiential goods have higher intrinsic information asymmetry, their consumption often relies on established trust signals, such as ratings and personal recommendations rather than aggregate buzz. In contrast, search goods are more prone to visibility-driven herd effects, as the decision to visit is primarily influenced by awareness rather than an assessment of quality [25]. We classify NightClub/Bar alongside Shopping Mall as search goods because the district-level visit decision—whether to go to Khaosan Road versus Thonglor—depends primarily on searchable attributes such as location awareness, crowd presence, and brand visibility [42,43]. Although within-venue nightlife quality is experienced only upon consumption, the spatial choice that our conditional logit captures operates at the pre-visit awareness stage of the decision hierarchy, where aggregate buzz directly informs the visit-or-not decision.

H3. *The buzz effect differs between experience goods (Spa, Restaurant) and search goods (Shopping, Nightlife), with search goods exhibiting stronger buzz sensitivity.*

The overtourism literature documents that popular districts generate disproportionately more user-generated content, creating a positive feedback loop between visibility and visitation [5,33]. If information cascades are naturally concentrated in areas already experiencing congestion, then social media buzz may, in practice, exacerbate existing spatial inequalities [3,4].

Because reinforcement is a causal claim that cannot be identified from a single cross-sectional correlation, we specify H4 as a claim about the causal visitation channel and test it with a panel-level instrumental-variable strategy in Section 4.5, rather than as a cross-sectional congestion correlation.

H4. *Social-media buzz causally increases district-day tourist visitation and thereby reinforces the spatial concentration of tourist flows toward high-buzz districts.*

Figure 1 summarizes the causal architecture implied by H1–H4. The top row traces a four-link chain from the social media information environment through the cascade-based behavioral mechanism to tourists’ spatial choices and ultimately to sustainability outcomes. The amber feedback loop at the bottom visualizes the self-reinforcing cascade whereby visitation concentration generates additional user-generated content and amplifies the cascade on subsequent days.

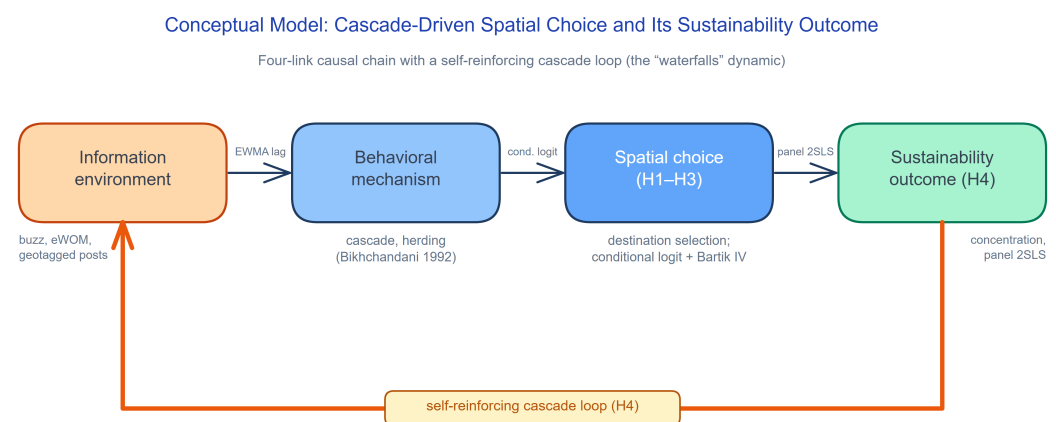


Figure 1. Conceptual model. Top row: causal chain from the social-media information environment through the cascade mechanism (following Bikhchandani et al. [19]) to spatial choice and the sustainability outcome. H1–H3 test the environment-to-choice links; H4 tests the choice-to-outcome link. The amber feedback loop visualizes cascade reinforcement.

3. Materials and Methods

3.1. Study Context of the Bangkok Songkran Festival

Songkran, the Thai New Year, is celebrated from April 11 to 17 annually and is characterized by citywide water fights, street parties, and religious ceremonies. During the 2019 Songkran period, Bangkok experienced a 28% increase in tourist taxi trips relative to the preceding week, creating significant spatial concentration challenges. Figure 2 displays the spatial distribution of social media buzz across Bangkok’s 48 administrative districts, illustrating the pronounced concentration of online discussion in a small number of central districts.

3.2. Data Sources

This study integrates three datasets covering the seven days before, during, and after Songkran, 4–24 April 2019.

Tourist-identified taxi trips were obtained from a larger dataset of 715,897 segmented taxi trajectories provided by the ITIC Foundation, from which 95,692 tourist trips were extracted. Each trip record contains origin and destination coordinates (latitude and longitude), timestamps, speed at endpoints, and semantic enrichment from the Google Places API, including point-of-interest (POI) name, category across seven types, and Google rating on a 1–5 scale. The dataset contained 6365 unique destination POIs.

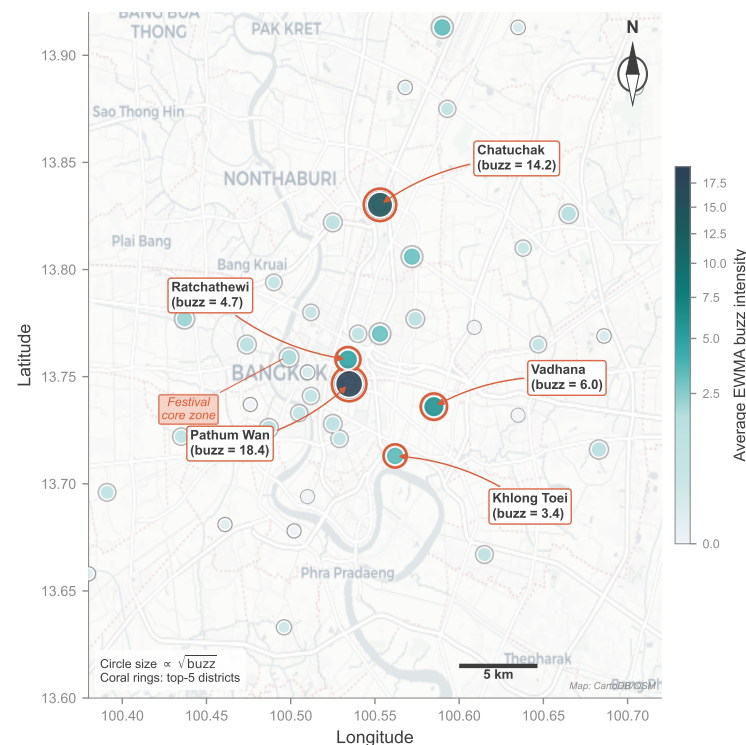


Figure 2. Spatial distribution of social media buzz intensity across 48 Bangkok administrative districts. The circle size is proportional to the average EWMA buzz value. The five highest-buzz districts (Pathum Wan, Chatuchak, Vadhana, Ratchathewi, and Khlong Toei) are labeled with their average values. The approximate festival core zone near the Old City is indicated; notably, buzz concentrates in modern commercial districts rather than the traditional festival area, consistent with the search-good sensitivity documented in H3.

Social media activity was captured through 5995 anonymized, geotagged Twitter posts collected over the same spatial and temporal coverage. Each record contained an anonymized user ID, timestamp (converted from UTC to Bangkok local time, GMT+7), administrative district, and activity type. Of these, 3216 records (53.6%) contain valid district information, spanning 59 district names, of which 48 are valid Bangkok administrative districts after cleaning. At the district–day resolution 73.5% of the 1008 cells record zero tweets, and the platform’s demographic composition and external-validity implications are discussed in Section 5.5.

Table 1 summarizes the key characteristics of each period. Tourist trip volume increased by 28% from the pre-festival to the festival period (3850 to 4930 daily trips), while social media buzz peaked during and immediately after Songkran, as visitors shared their experiences online. The number of active districts remained constant at 48, indicating that the spatial footprint of tourism did not contract during the festival, even as the volume surged.

Drawing on the full taxi fleet (334,312 cleaned trips), we computed a grid-level congestion index for 14,641 spatial grid cells ($\sim 500\text{ m} \times 500\text{ m}$) at 2-h intervals as follows:

$$CI_{jt} = 1 - \frac{\bar{v}_{jt}}{v_j^{\text{free-flow}}} \quad (1)$$

where $\bar{v}_{jt}$ is the average taxi speed in grid cell j at time t , and $v_j^{\text{free-flow}}$ is the 90th percentile speed during off-peak hours (22:00–06:00). The mean congestion index was 0.545 (median 0.562). Although this vehicle-speed-based metric does not directly capture pedestrian-level crowding, in a dense urban environment such as Bangkok, where road congestion and

foot traffic are strongly correlated owing to shared commercial corridors, limited sidewalk capacity, and the prevalence of street-level commercial activity, taxi fleet speed serves as an effective proxy for overall area-level activity intensity and physical congestion [13]. Note that the proxy does not capture pedestrian crowding on walking streets such as Khaosan Road during Songkran, where vehicle access is restricted. Mobile-phone-signal or on-site-survey measurements would provide complementary pedestrian-scale coverage [41,44].

Table 1. Descriptive statistics by study period.

Period	Dates	Daily Trips	Daily Buzz (District-Summed EWMA)	Mean Gini	Districts
Before	4–10 April	3850	14.1	0.541	48
During	11–17 April	4930	104.2	0.529	48
After	18–24 April	4890	136.6	0.517	48

Daily Trips and Daily Buzz are period averages. Daily Buzz reports $\sum_{d=1}^{48} \text{Buzz}_{d,t}$, the sum of district-level EWMA buzz across all 48 districts ($\lambda = 0.5$, $L = 3$). Gini coefficient measures spatial concentration of tourist trips across 48 districts.

3.3. Social Media Buzz Variable Construction

To provide spatial variation across choice alternatives, we mapped each of the 6365 POIs to its nearest Bangkok administrative district using Euclidean distance to district centroids, which yielded 48 districts with at least one POI. The spatial imprecision of nearest-centroid assignment for POIs close to administrative boundaries is assessed by the buffer robustness reported in Table A1 (Panel A).

Raw daily tweet counts per district exhibit severe zero-inflation, with 73.5% of district-day combinations recording zero tweets. To address this while maintaining the temporal lag structure required for causal identification, we construct the following EWMA buzz variable:

$$\text{Buzz}_{d,t} = \sum_{s=t-L}^{t-1} e^{-\lambda(t-s)} \cdot n_{d,s}, \quad L \in \{1, 2, 3, 5\}, \lambda \in \{0.3, 0.5, 0.7, 1.0\}. \quad (2)$$

where $n_{d,s}$ is the tweet count in district d on day s , L is the lookback window, and λ is the decay rate. The summation bounds run from $s = t - L$ through $s = t - 1$, so the current-day tweet volume is excluded from the regressor, and the temporal priority of information relative to choice is enforced. The default specification uses $L = 3$ days and $\lambda = 0.5$. This construction ensures that only past social media activity informs the buzz variable, ruling out reverse causality on the same day. The baseline calibration $\lambda = 0.5$ corresponds to a decay half-life of $\ln 2 / 0.5 \approx 1.4$ days, which matches the empirical autocorrelation length of district-level tweet volume in our window (lag-1 $\rho \approx 0.5$ decaying to noise by lag-3) and is consistent with the 24–72 h engagement half-life documented for short-form social-media content. Slower decays such as $\lambda = 0.3$ overweight stale activity that is unlikely to remain salient at the choice moment, while faster decays such as $\lambda = 1.0$ collapse the window onto the previous day and discard the multi-day word-of-mouth dynamics that the cascade mechanism is meant to capture. The 4×4 parameter grid in Table 2 confirms that the substantive conclusions are invariant to this choice across all 32 configurations, with both the buzz main effect and the festival interaction remaining significant at $p < 0.001$ throughout.

Table 2. Sensitivity of buzz and festival interaction coefficients to EWMA parameters.

Lookback	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 1.0$
<i>Panel A: β_{Buzz}</i>				
$L = 1$ day	+0.0105	+0.0105	+0.0105	+0.0105
$L = 2$ days	+0.0115	+0.0115	+0.0115	+0.0113
$L = 3$ days	+0.0125	+0.0124	+0.0122	+0.0119
$L = 5$ days	+0.0133	+0.0131	+0.0127	+0.0122
<i>Panel B: $\gamma_{\text{During} \times \text{Buzz}}$</i>				
$L = 1$ day	+0.0037	+0.0037	+0.0037	+0.0037
$L = 2$ days	+0.0043	+0.0042	+0.0041	+0.0040
$L = 3$ days	+0.0046	+0.0043	+0.0042	+0.0040
$L = 5$ days	+0.0075	+0.0059	+0.0050	+0.0043

All 32 coefficients are statistically significant at $p < 0.001$. Panel A reports the buzz main effect β_{Buzz} ; Panel B reports the festival interaction $\gamma_{\text{During} \times \text{Buzz}}$.

3.4. Choice Set Construction

For each observed tourist trip, we constructed a discrete choice set comprising the actually chosen destination POI alongside $K = 20$ unchosen alternative POIs sampled from within a spatial buffer of the trip origin. The buffer radius of 11.2 km corresponds to the 90th percentile of observed trip distances, ensuring that the consideration set reflects realistic travel ranges. To guard against sampling bias, we adopted a stratified scheme rather than simple random selection: two alternatives were drawn from the same POI category as the chosen destination, two from the nearest-distance quintile, and the remaining 16 at random. This procedure generated 2,009,450 choice records across 95,692 trips, with each trip presenting the tourist with 21 competing alternatives. Consistent estimation of conditional-logit parameters under sampled alternatives rests on McFadden [45], who established the sampling-of-alternatives result under Independence of Irrelevant Alternatives and a uniform conditioning rule. Train [46] shows in Chapter 3 that K in the range from 10 to 30 yields parameter estimates indistinguishable from full-choice-set estimation at a modest efficiency cost. Nerella and Bhat [47] reported analogous results for spatial destination choice. Table A1 (Panel B) reports the main buzz coefficient across ten configurations formed by crossing $K \in \{10, 20, 30\}$ with distance-percentile thresholds $P \in \{75, 85, 90, 95\}$, with the coefficient stable in the narrow band $[+0.0097, +0.0109]$ at $p < 0.001$ in every configuration.

3.5. Utility Specification

Under the random utility maximization (RUM) framework, tourist n derives utility U_{njt} from choosing destination j at time t :

$$U_{njt} = V_{njt} + \varepsilon_{njt} \quad (3)$$

where V_{njt} is the systematic utility and ε_{njt} is an i.i.d. Type I Extreme Value error. The systematic utility is specified as

$$\begin{aligned} V_{njt} = & \beta_1 \cdot \text{TravelTime}_{njt} + \beta_2 \cdot \text{Rating}_j + \beta_3 \cdot \text{CompDensity}_j \\ & + \beta_4 \cdot \text{AggloDensity}_j + \beta_5 \cdot \text{CI}_{jt} + \beta_6 \cdot \text{Buzz}_{d(j),t} \\ & + \gamma_1(\text{During}_t \times \text{TravelTime}_{njt}) + \gamma_2(\text{During}_t \times \text{Rating}_j) \\ & + \gamma_3(\text{During}_t \times \text{Buzz}_{d(j),t}) \end{aligned} \quad (4)$$

Table 3 summarizes all the variables.

The model was estimated using the maximum likelihood method. To extend the analysis, we also estimate a model with an experience goods interaction term ($\text{ExpGoods}_j \times \text{Buzz}_{d(j),t}$), where experience goods are defined as Spa and Restaurant categories.

Table 3. Variable definitions and descriptive statistics.

Variable	Description	Source	Mean	SD
TravelTime	Estimated travel time (min)	GPS	6.72	5.81
Rating	Google rating (1–5)	Google Places	3.33	1.21
CompDensity	Same-category POIs within 500 m	Spatial analysis	4.50	5.89
AggloDensity	Different-category POIs within 500 m	Spatial analysis	11.40	11.73
CI_{jt}	Congestion index (0–1)	Full taxi fleet	0.55	0.27
$\text{Buzz}_{d(j),t}$	EWMA social media buzz	Twitter	4.21	7.35
During_t	Festival period dummy (11–17 April)	Calendar	0.36	—

3.6. Identification Strategy

Three features of our research design jointly address these endogeneity concerns. The most fundamental is the temporal lag structure: because buzz is measured using tweets from days $t - 1$ to $t - 3$ while destination choices occur on day t , same-day reverse causality is ruled out by construction. This temporal separation is reinforced by spatial aggregation asymmetry, whereby buzz is measured at the district level, while choices are observed at the finer-grained POI level, reducing the mechanical correlation between the dependent and independent variables. Finally, we implement a placebo test in which buzz values are randomly permuted across districts, and the model is re-estimated 50 times. If the real buzz coefficient falls outside the 95% confidence interval of this permutation distribution, the observed effect can be attributed to genuine information influence rather than spatial autocorrelation or unobserved confounders.

To address the residual concern that time-varying district-specific unobservables may drive both social-media buzz and tourist visitation, we construct a Bartik shift-share instrument [20,21]:

$$Z_{d,t} = s_d \cdot G_{-d,t}, \quad (5)$$

where s_d is a district-level share and $G_{-d,t}$ is the platform-wide aggregate tweet count on day t excluding district d 's own tweets through a leave-one-out construction. We evaluate three candidate share constructions: the All-POI-share (primary), which takes s_d as each district's proportion of all POI categories; the Tweet-share, based on each district's historical share of total Twitter activity; and the Buzz-POI-share (original), based on the proportion of Shopping Mall and NightClub/Bar POIs. Because the conditional-logit model is non-linear, we implement the IV through a control-function approach [48,49], first estimating

$$\text{Buzz}_{d,t} = \alpha + \delta \cdot Z_{d,t} + \mu_d + \varepsilon_{d,t}, \quad (6)$$

and inserting the residuals as an additional regressor in the second-stage conditional logit. The Rotemberg-weight decomposition, the Borusyak–Hull shift-orthogonality test, and the three-IV cross-validation are reported alongside the results in Section 4.4.

3.7. Sustainability Assessment

To assess whether social media buzz facilitates or hinders sustainable spatial distribution, we computed two complementary concentration metrics. The Gini coefficient, applied to daily tourist trip counts across districts, captures the degree of spatial inequality in the tourist distribution. The Herfindahl–Hirschman Index (HHI), computed from district-level trip shares, provides an alternative measure of market-style concentration.

Together, these indices track whether the spatial distribution of tourists becomes more or less equitable over the study period. We further examine the cross-sectional correlation between district-level buzz intensity and congestion to determine whether information cascades reinforce or alleviate existing spatial imbalances. To test whether the observed Gini trajectory represents a statistically meaningful change rather than a day-to-day fluctuation, we apply parametric and non-parametric trend tests together with period-level ANOVA, Kruskal–Wallis, and bootstrap-based mean-difference tests to the 21-day district-level Gini series; full results are reported in Table A4 (Panel A).

4. Results

4.1. Main Model Results

The conditional logit model confirms that social media buzz exerts a significant and substantive effect on tourists' spatial choices (Table 4).

Table 4. Conditional logit results: Main model and extended model with experience goods interaction.

Variable	Main Model		Extended Model (H3)	
	Estimate	z-Stat	Estimate	z-Stat
TravelTime	−0.048 ***	−150	−0.048 ***	−236
Rating	−0.069 ***	−27.8	−0.068 ***	−34.5
CompDensity	−0.107 ***	−116	−0.107 ***	−133
AggloDensity	+0.035 ***	+140	+0.034 ***	+161
Congestion	+1.275 ***	+53.6	+1.274 ***	+54.6
Buzz	+0.010 ***	+18.2	+0.007 ***	+15.1
During × TravelTime	+0.022 ***	+53.1	+0.022 ***	+83.8
During × Rating	+0.018 ***	+4.30	+0.017 ***	+5.16
During × Buzz	+0.003 **	+2.97	+0.004 ***	+4.94
ExpGoods × Buzz	—	—	−0.003 ***	−4.90
Log-likelihood	−258,807		−258,974	
McFadden ρ^2	0.1116		0.1111	
Parameters	9		10	

N = 95,692 trips; 2,009,450 choice records. *** $p < 0.001$, ** $p < 0.01$. Standard errors computed via numerical Hessian.

Base utility parameters carry the expected signs: tourists favor proximate destinations with diverse POI clusters and avoid areas of intense same-category competition (see Figure 3 for the complete coefficient profile). Two coefficients merit brief comment.

The negative pooled Rating coefficient reflects category-specific heterogeneity rather than misspecification. POI-category fixed effects reverse the sign to positive (Table A3, Panel A), and subsample regressions (Table A3, Panel B) show that Restaurant and Shopping Mall visitors prefer higher-rated venues, while Spa and NightClub/Bar visitors prefer lower-rated venues, a pattern consistent with authenticity-seeking in categories where high ratings may signal touristic over-commercialization. Because Spa and NightClub/Bar trips dominate the sample, the pooled coefficient inherits their negative sign. The positive Congestion coefficient similarly proxies area-level vibrancy and is essentially unaffected by the addition of category fixed effects.

The central finding is that district-level social media buzz significantly increases destination choice probability ($\beta = +0.010$, $p < 0.001$), supporting H1. A district at Pathum Wan's buzz intensity enjoys roughly 20% higher odds of being chosen than a district with negligible social media presence, an advantage equivalent to approximately four minutes of travel-time reduction. Social media buzz, in effect, makes a district feel closer than it actually is.

Consistent with H2, the festival intensified this cascade. The marginal buzz effect rises by roughly 27% during Songkran, elevating the Pathum Wan odds advantage accordingly (Figure 3). The concurrent reduction in travel time sensitivity reinforces this interpretation: when private signals become noisier through road closures, crowd dynamics, and altered schedules, tourists become simultaneously more responsive to digital signals and more willing to venture beyond their immediate vicinity.

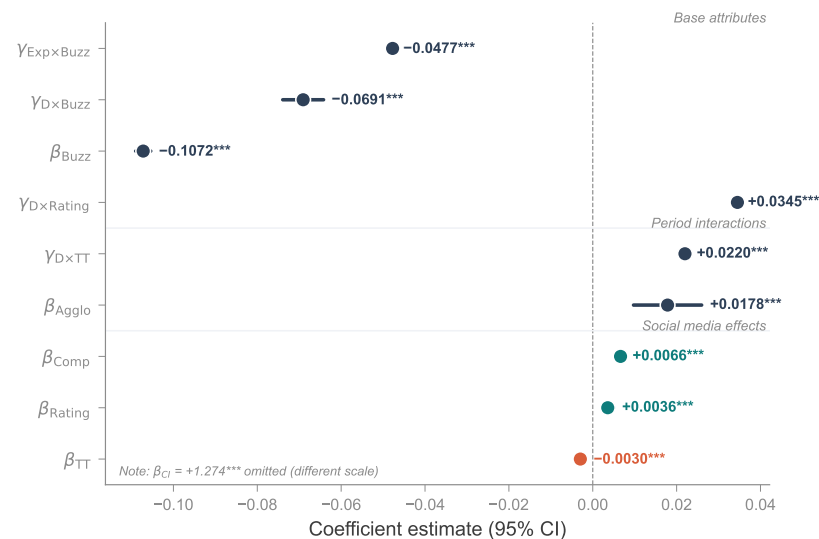


Figure 3. Forest plot of conditional logit coefficient estimates with 95% confidence intervals (extended model with experience goods interaction). The congestion index coefficient ($\beta_{CI} = +1.274$ ***) is omitted because of the different scale. All displayed coefficients are significant at $p < 0.001$ (denoted ***).

In contrast, not all destination types are equally susceptible to cascade dynamics (H3). Experience goods, such as spas and restaurants, are significantly less responsive to district-level buzz than search goods, such as shopping malls and nightlife venues. This asymmetry maps directly onto the classical search-versus-experience distinction in the economics of information [42,43]: search-good quality can be evaluated from publicly observable attributes prior to consumption, so visibility-driven cascades dominate the awareness problem that drives the choice; experience-good quality can only be assessed through consumption, so private trust signals such as personal recommendations and within-platform ratings outweigh aggregate buzz [25]. The monotone gradient observed in the four-category specification of Table A3 (Panel B), running from pure-search Shopping Mall ($\beta_{Buzz} = +0.0210$) and NightClub/Bar (+0.0199) through experience-leaning Spa (+0.0073) to pure-experience Restaurant (+0.0040), traces this Nelson–Klein continuum precisely and converts what would otherwise be a post hoc empirical pattern into a theoretically anchored prediction of differential cascade sensitivity.

4.2. Placebo Test

Spatial confounding was ruled out by a permutation placebo in which district-level buzz values were randomly reassigned, and the model was re-estimated 50 times. The real buzz coefficient exceeds the placebo upper bound by roughly an order of magnitude, with a permutation p -value well below conventional thresholds (Figure 4; Table 5), confirming that the observed buzz–choice relationship reflects genuine information influence rather than unobserved district-level confounders.

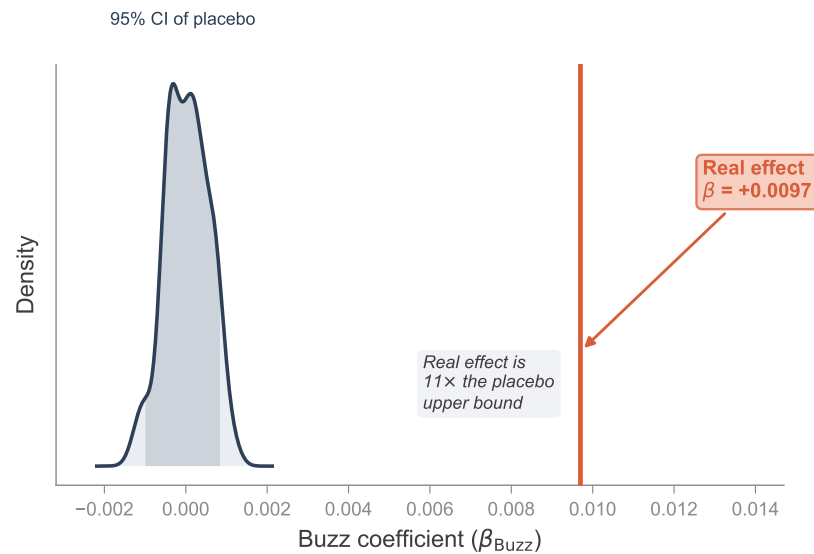


Figure 4. Placebo test results. The real buzz coefficient (+0.010, vertical line) falls far outside the 95% confidence interval of the placebo distribution. The real effect is approximately $11\times$ larger than the placebo upper bound, confirming that the buzz–choice relationship is genuine.

Table 5. Placebo test summary statistics.

Statistic	Value
Real β_{Buzz}	+0.0097
Placebo mean (50 permutations)	+0.0001
Placebo SD	0.0005
Placebo 95% CI	[−0.0010, +0.0009]
Real/Placebo upper bound	$\sim 11\times$
Permutation p -value	< 0.02

Placebo coefficients obtained by randomly permuting district-level buzz values across districts 50 times and re-estimating the conditional logit model. The permutation p -value is computed as the proportion of placebo coefficients exceeding the real coefficient.

4.3. Sensitivity Analysis

The findings survived three sensitivity tests. Re-estimating the model across a 4×4 grid of the EWMA lookback windows and decay rates yields uniformly positive and significant coefficients for both the main buzz effect and the festival interaction (Table 2). A clear gradient emerges along which longer lookback windows and slower decay rates systematically strengthen both effects (Figure 5). For destination managers, this gradient implies that sustained content campaigns will outperform isolated promotional campaigns.

The main buzz coefficient is similarly robust to the choice-set construction rule. Across ten alternative configurations formed by crossing the number of non-chosen alternatives $K \in \{10, 20, 30\}$ with the travel distance percentile threshold $P \in \{75, 85, 90, 95\}$, the estimate remains in a narrow band with no loss of significance (Table A1, Panel B). Our baseline $K = 20$ choice rests on the sampled-alternatives consistency result of McFadden [45], the operational guidance in Train [46], and the spatial-destination-choice analog reported by Nerella and Bhat [47].

We finally probe the spatial imprecision of nearest-centroid district assignment by progressively dropping POIs whose second-nearest centroid lies within a 0.5-, 1-, or 2 km buffer of the nearest centroid district. The buzz coefficient is stable under the 0.5- and 1-km filters and only weakens under the aggressive 2-km filter, which retains roughly one-fifth of the sample (Table A1, Panel A). The slightly larger coefficients under the moderate filters are consistent with a mild measurement-attenuation bias in the baseline when boundary-

ambiguous POIs are included, which reinforces rather than undermines the main finding.

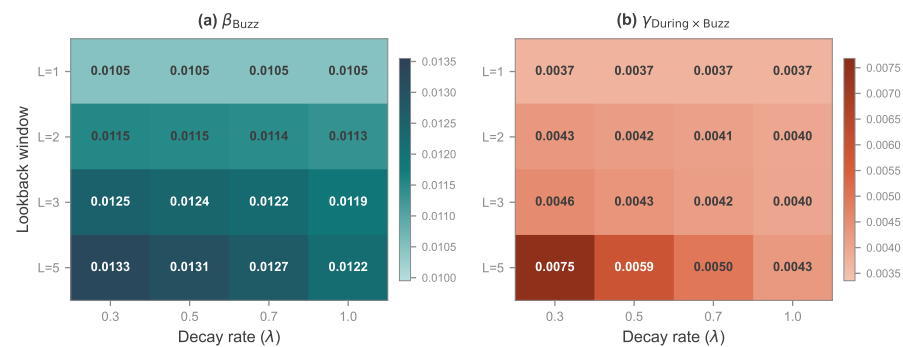


Figure 5. Sensitivity of EWMA parameter choices. (a) Buzz main effect (β_{Buzz}) across 16 lookback $\times$ decay rate configurations. (b) Festival interaction ($\gamma_{\text{During} \times \text{Buzz}}$). All 32 coefficients are significant at $p < 0.001$. Darker cells indicate larger coefficients.

4.4. Bartik Shift-Share Instrumental Variable Estimation

Applying the Bartik shift-share IV strategy specified in Section 3.6, we first assess the identifying assumption through Rotemberg, cross-IV, and Borusyak–Hull diagnostics (Table A2). The All-POI-share produces the most diffuse Rotemberg-weight distribution, with the top weight falling on an outer commercial-residential district that is not a high-buzz destination, whereas the Tweet-share and Buzz-POI-share each concentrate identification on a single high-buzz district. Therefore, we adopt All-POI-share as the primary specification. The three IVs yield coefficients within a band of under 5%, mitigating the concern that any single share construction drives the result. The Borusyak–Hull test returns a positive residual–shift correlation that we attribute to the Songkran common shock; identification therefore rests on the combined strength of the temporal lag, the placebo permutation, three-IV convergence, and diffuse Rotemberg weights rather than on any single diagnostic.

We then estimate the Bartik IV separately for each of the three study periods rather than using a single pooled interaction model, so that heterogeneous instrument relevance across periods is accommodated directly and H2 can be tested as a causal cross-period comparison (Table 6).

The instrument achieves strong first-stage relevance during and after the festival but fails in the pre-festival window, where near-zero Twitter activity deprives it of temporal leverage [50]. Therefore, the period before the study was excluded from the IV inference as a conservative choice reflecting data sparsity rather than selective reporting.

Two substantive findings were obtained. First, the IV-corrected festival coefficient exceeds the OLS estimate by 51%, a pattern diagnostic of classical measurement-error attenuation: sparse, noisy Twitter data bias the OLS coefficient toward zero, and once the control function purges this attenuation the true causal effect emerges as substantially stronger. In practical terms, a district at Pathum Wan’s buzz level now commands 42% higher odds of being chosen compared with 20% under OLS. Second, the During coefficient exceeds the After coefficient by 61% (Wald $z = 6.05$, $p < 0.001$), confirming the festival-amplification hypothesis (H2) in its causal form. This cross-period amplification exceeds the 27% implied by the pooled interaction model, indicating that pooling across periods with heterogeneous instrument strength understates the true festival-driven intensification.

Taken together, the three identification strategies converge: the temporal lag rules out same-day reverse causality, the placebo test rules out spatial confounding, and the Bartik IV rules out time-varying district-level unobservables. This triangulation provides robust causal evidence that the intensity of social media discussions drives tourists’ spatial choices.

Table 6. Split-period Bartik IV estimation results.

Period	First Stage		OLS (No IV)		Control Function	
	<i>F</i> -Stat	<i>R</i> ²	β_{Buzz}	<i>p</i>	$\beta_{\text{Buzz}}^{\text{IV}}$	<i>p</i>
Before (4–10 April)	0.08	0.413	+0.068 ***	<0.001	excluded (weak instrument)	
During (11–17 April)	90.59 ***	0.568	+0.012 ***	<0.001	+0.019 ***	<0.001
After (18–24 April)	9.64 **	0.844	+0.010 ***	<0.001	+0.012 ***	<0.001
Wald test: $\beta_{\text{During}}^{\text{IV}} - \beta_{\text{After}}^{\text{IV}}$					+0.007	<0.001
Amplification ratio: During/After						1.61 ×

$N_{\text{Before}} = 26,948$; $N_{\text{During}} = 34,513$; $N_{\text{After}} = 34,231$ trips. First-stage *F*-statistics test the null that the Bartik instrument is irrelevant; the conventional threshold for instrument strength is $F > 10$. The Before period is excluded from IV inference because $F = 0.08$ indicates a weak instrument, attributable to the near absence of Twitter activity prior to the festival. Control function residual coefficients are significant at $p < 0.001$ in both the During and After periods, confirming the presence of endogeneity in the original buzz variable. *** $p < 0.001$, ** $p < 0.01$.

4.5. Hotspot Reinforcement and Spatial Sustainability (H4)

We test H4 at two levels of aggregation. Across districts, buzz and congestion were positively correlated ($r = +0.38$, $p = 0.011$); however, this correlation alone conflates within-district buzz dynamics with pre-existing differences in district attractiveness. Therefore, we estimate two panel-level IV specifications using the All-POI-share Bartik IV with district and date fixed effects. Regressing district-day congestion on buzz yields a null causal coefficient, indicating that the district-level correlation is driven principally by these attractiveness differences rather than a direct within-district channel. In contrast, regressing district-day tourist visitation on buzz yields a positive and significant causal effect of approximately 4.7 extra trips per unit of buzz (Table A4, Panel B). H4 is therefore supported in causal form through the tourist-flow channel, and the null vehicular-congestion effect reflects the modest share of tourist trips in total roadway traffic rather than the absence of the cascade mechanism.

All five top-buzz districts exhibit above-average congestion, and the pattern persists through the top ten (Table 7). Notably, the two districts that bookend the buzz distribution are both dominated by search-good POIs, consistent with the finding that search goods exhibit stronger buzz sensitivity (H3). The convergence of high buzz, high congestion, and search-good dominance suggests that information cascades channel tourists disproportionately toward already strained entertainment and shopping corridors [5,33].

At the aggregate city level, spatial concentration declines monotonically over the study period (Table 8): the Top 3 district share falls from 27.6% to 25.8%, redistributing roughly 1800 daily trips toward peripheral areas. Five complementary tests (Table A4, Panel A) confirm the decline is statistically significant, with parametric and non-parametric trend tests as well as the three-period ANOVA, Kruskal–Wallis, and a bootstrap-based Before-versus-After difference all rejecting the null. The weaker daily buzz–Gini correlation ($r = -0.34$) addresses a narrower day-to-day question whose null should be scoped to that specific claim rather than read as contradicting the monotone trend.

Table 7. Top 10 districts by social media buzz intensity.

Rank	District	Avg Buzz	Avg CI	Dominant Category	Search %
1	Pathum Wan	18.39	0.656	NightClub/Bar	67
2	Chatuchak	14.22	0.617	Restaurant	38
3	Vadhana	5.95	0.635	Restaurant	45
4	Ratchathewi	4.65	0.655	Restaurant	35

Table 7. Cont.

Rank	District	Avg Buzz	Avg CI	Dominant Category	Search %
5	Khlong Toei	3.36	0.631	Restaurant	31
6	Don Mueang	3.24	0.653	Restaurant	30
7	Din Daeng	3.04	0.653	Restaurant	39
8	Lat Phrao	3.02	0.619	Restaurant	39
9	Taling Chan	2.14	0.608	Restaurant	28
10	Phra Nakhon	1.89	0.631	NightClub/Bar	40

Avg Buzz: mean EWMA buzz intensity ($\lambda = 0.5$, $L = 3$). Avg CI: mean congestion index. Search %: share of trips to search for good POIs (Shopping Mall + NightClub/Bar). Pathum Wan and Phra Nakhon are the only NightClub/Bar-dominant districts, consistent with the H3 finding that search goods exhibit stronger buzz sensitivity. Pearson $r = +0.38$, $p = 0.011$ ($N = 43$ districts).

Figure 6 synthesizes the sustainability evidence across the four panels.

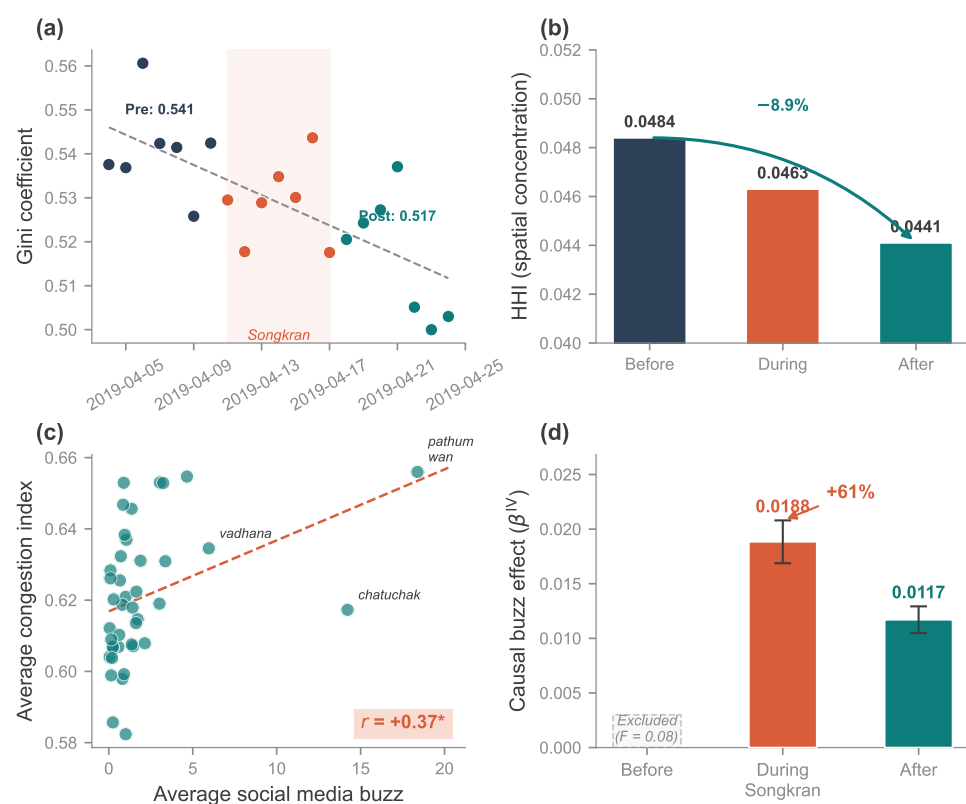


Figure 6. Sustainability assessment dashboard. (a) Daily Gini coefficient of tourist spatial concentration, with the Songkran festival period shaded coral; the dashed grey line shows the linear trend of daily Gini values. (b) Herfindahl–Hirschman Index by period; the curved arrow indicates the period-over-period decline in HHI (−8.9%). (c) District-level scatter of average buzz versus average congestion index, with the top districts labeled; the dashed red line is the OLS fit and * denotes $p < 0.05$ for the Pearson correlation coefficient ($r = +0.37$). (d) IV-corrected causal buzz effect from the split-period Bartik estimation; the Before period is excluded for weak-instrument reasons; the vertical arrow indicates the +61% change in the IV-corrected causal buzz effect between the During-Songkran and After periods. Error bars represent 95% CIs.

Table 8. Spatial concentration of tourist trips by period.

Period	Gini	HHI	Top-3 Share	Active Districts
Before (4–10 April)	0.541	0.048	27.6%	48
During (11–17 April)	0.529	0.046	26.6%	48
After (18–24 April)	0.517	0.044	25.8%	48

5. Discussion

5.1. Social Media as a Double-Edged Sword for Sustainable Tourism

Our findings reveal a paradox at the heart of the relationship between digital information and sustainable tourist mobility. At the district level, unmanaged cascades reinforce existing spatial hotspots, consistent with the self-reinforcing dynamics predicted by cascade theory [19,33]. At the city level, however, aggregate spatial concentration declines significantly over the study window. The two trends together suggest that information cascades neither uniformly concentrate nor uniformly disperse tourists; they operate through two countervailing channels whose outcomes differ by spatial scale. This scale-dependent ambiguity is what makes proactive governance of the digital information environment essential because, in its absence, the hotspot-reinforcing channel is likely to prevail.

The sustainability consequences of this tension extend across all three pillars of sustainable development.

Environmentally, the highest-buzz districts are also the most traffic-congested, with average vehicle speeds falling well below the free-flow levels. Because fuel consumption and CO₂ emissions per vehicle-kilometer rise non-linearly with congestion severity [51], channeling a disproportionate share of tourist trips into a handful of corridors generates outsized transport emissions relative to those produced by a more dispersed pattern. Beyond motorized transport, concentrated foot traffic in peak districts accelerates wear on public spaces, strains waste collection, and raises ambient noise in mixed-use neighborhoods. Steering tourists into areas where marginal environmental damage per additional visitor is the highest runs counter to the SDG 12 aspiration that information systems should support sustainable consumption.

Socially, reinforcing tourist flows in mixed-use historic districts has direct consequences for residents' well-being. In Bangkok's Old City, where major heritage sites sit alongside dense residential communities, heightened congestion reflects the daily friction between visitor activity and the routines of residents who share the same narrow streets and limited transit infrastructure. Nightlife-dominated commercial districts also impose late-night noise, waste, and safety externalities on adjacent residential areas. Where cascades intensify tourist concentration in these contested zones, they amplify overtourism pressures that erode urban livability [44] and work against the SDG 11 goal of reducing cities' adverse environmental impacts.

Spatial concentration refers to revenue concentration. When a small set of districts captures a disproportionate share of tourist trips, peripheral communities receive almost no expenditure and are effectively excluded from the festival's economy. The post-festival shift toward a more dispersed distribution therefore represents a meaningful reallocation of tourism revenue, allowing peripheral districts to develop their own tourism economies rather than serving as bedrooms for a centrally located festival. The strategic governance of the digital information environment can accelerate this redistribution by increasing the visibility of under-visited areas.

Two interpretive caveats deserve explicit treatment in this discussion rather than relegation to the methodological footnotes, because they bear directly on how readers should calibrate the causal claim. The first concerns the distinctive traffic-management context of the Songkran festival. The festival is accompanied by extensive city-managed restrictions, including official road closures around Sanam Luang and the Grand Palace, pedestrian-only zones along Khaosan Road, and rerouting of public transport, all of which depress taxi-derived vehicle speeds independently of any social-media-driven behavioral pull. Three features of our identification strategy disentangle these management-induced effects from the cascade-driven attraction that the model is designed to measure. The panel 2SLS specification with district and date fixed effects (Table A4, Panel B) absorbs city-wide

festival-day shocks that act on all districts simultaneously, including any aggregate component of management-induced traffic congestion. The IV-identified visitation effect of +4.67 extra trips per unit of buzz operates on tourist flow rather than on vehicle speed, and within-district within-day buzz movements do not produce a detectable causal effect on the congestion proxy itself ($\hat{\beta}^{IV} = -0.0025$, $p = 0.155$); therefore, the cross-sectional buzz–congestion correlation that originally motivated H4 is best read as the joint footprint of cascade-driven visitation and structural cross-district attractiveness rather than as evidence that buzz mechanically slows traffic. Because management-induced restrictions are concentrated in a small set of historic district corridors and applied uniformly during festival hours, their footprint is captured by the date and district fixed effects rather than by within-district variation in social media buzz. Future work that combines taxi GPS with pedestrian-flow sensors or BTS tap-in/tap-out data would allow a cleaner decomposition of vehicular versus pedestrian congestion sources.

The second caveat concerns whether the IV-identified causal effect is artifactually driven by super-hotspots such as Pathum Wan, whose buzz volume is roughly four times the city average. The Rotemberg-weight diagnostics in Table A2 directly address this issue. Under the primary All-POI-share instrument, the largest district weight is 0.317 for Bang Na, an outer commercial–residential district that does not appear among the ten highest-buzz destinations, and the effective number of identifying districts is 7.4. The Tweet-share variant, which by construction loads heavily on Pathum Wan (top weight 0.661), still yields an IV coefficient of +0.0215, which is statistically indistinguishable from the All-POI-share estimate of +0.0226 and from the Buzz-POI-share estimate of +0.0218. The convergence across instruments built on structurally different identifying districts indicates that the causal buzz effect is shared across quieter and busier districts rather than concentrated on any single high-buzz corridor. This consistency is what allows us to interpret the cascade as a city-wide behavioral mechanism rather than as the empirical signature of one famous shopping district.

However, the robustness of the cascade mechanism also makes strategic interventions viable. The IV-corrected estimate shows that the true behavioral response to social media is substantially stronger than conventional estimates suggest, which implies that DMO content interventions can generate meaningful spatial redistribution. Because this causal effect is strongest in festival windows, such interventions would deliver the largest gains precisely when overtourism pressure is most severe.

5.2. Implications for Information-Based Destination Management

This is an observational study, not an evaluation of a designed intervention such as a digital nudge [6]; what it identifies is the causal behavioral substrate on which any information-based visitor flow management strategy must rest. The implications below should therefore be read as calibrating guidance for DMOs weighing information-based content strategies rather than as validation of a specific intervention class.

Festival amplification of the cascade effect aligns sustainability needs with intervention leverage in a manner that is both theoretically informative and practically exploitable. Tourist volumes surge during festivals, intensifying congestion, waste generation, and resident disruption [52], yet our results show that social media responsiveness also peaks under these conditions. Information-based content programs need not operate year-round, then, because concentrated efforts during predictable high-stress windows can deliver disproportionate gains in spatial equity. The concurrent drop in travel-time sensitivity during festivals reinforces this point: tourists become more receptive to digital signals and more willing to venture beyond their immediate vicinity, opening a behavioral window in which redistribution toward peripheral areas meets the least resistance.

The category moderation finding adds a second dimension to the intervention design. Search-good destinations are substantially more buzz-sensitive than experience-good destinations, so content strategies aimed at under-visited shopping or nightlife districts will generate the largest redistribution response, while redirecting spa or restaurant traffic calls for different instruments such as curated recommendation lists or influencer partnerships. The within-category Rating heterogeneity documented in Section 4.1 further qualifies this guidance: in experience-good categories where high ratings may signal touristic over-commercialization rather than quality, official rating-boosting strategies risk backfiring. This type of differentiation avoids the inefficiency of uniform campaigns and concentrates marketing resources where the sustainability return is highest.

5.3. Theoretical Contributions

This study advances the theoretical literature in four ways. First, it carries information cascade theory [19] beyond its traditional domains in financial markets and technology adoption into the spatial tourism setting, showing that cascade dynamics, once confined to sequential decision contexts with binary outcomes, also operate in continuous spatial choice settings, where tourists select among geographically distributed alternatives.

Second, it enriches the eWOM literature [15,25] by establishing that area-level social media intensity, not merely establishment-level reviews, shapes spatial destination choices. Shifting the unit of analysis from the individual firm to the urban district opens a new line of spatially informed eWOM research that treats the information environment as a geographic field rather than a collection of discrete product evaluations.

Third, the study identifies two empirical boundary conditions for information-based visitor flow management. Situational uncertainty in the form of festival windows amplifies the causal buzz effect. Category heterogeneity drives search goods to respond more strongly than experience goods, with Spa and NightClub/Bar visitors within the experience-good set preferring lower-rated venues. Both boundary conditions have long been proposed conceptually but not empirically established.

Fourth, Bartik shift-share IV estimation [20,21] pushes the causal evidence beyond what temporal ordering and placebo tests can deliver. That the IV-corrected effect substantially exceeds the conventional estimate shows that measurement error in sparse social media data systematically attenuates observed effects, a methodological insight with broad implications for the eWOM literature, where social media variables are routinely measured with noise. A triple diagnostic combining Rotemberg-weight decomposition, the Borusyak–Hull test, and panel 2SLS gives future tourism eWOM studies a workable template for credible causal identification.

Beyond these discipline-specific contributions, the study speaks to the broader “twin transition” discourse in sustainable tourism [7,8]. The prevailing thesis is that digitalization and environmental sustainability advance in tandem [53], yet the empirical literature has largely treated them as complementary, while paying limited attention to the tensions between them. Our findings complicate this optimistic reading because the same digital platforms that enable scalable information-based visitor management simultaneously generate organic cascades that reinforce unsustainable spatial concentration. In other words, digitalization is not inherently sustainability-enhancing; its net contribution depends on whether the digital information environment is actively governed or left to amplify market-driven concentration. This conditional view provides a more nuanced theoretical foundation for twin transitions than the standard assumption of mutual reinforcement in the literature.

5.4. Practical Implications for Sustainable Tourism Management

For DMOs and urban planners, our findings suggest a three-stage adaptive management cycle aligned with the twin transition agenda [7,8] that includes: The first stage, monitoring, tracks real-time social media buzz by district to flag emerging hotspots before physical overcrowding materializes, enabling pre-emptive rather than reactive management, as follows: The second stage, information-based content governance, proactively generates engaging digital content for under-visited districts with carrying capacity. Because search-good categories are the most buzz-sensitive, campaigns should prioritize these destination types through influencer partnerships and curated local experience features. The third stage, evaluation, closes the loop by using GPS trajectory data to test whether newly governed content actually shifts tourist flows, providing an evidence-based feedback mechanism for the first stage.

Taken as a whole, this monitor–govern–evaluate cycle turns social media from a passive information channel into an active governance instrument for spatial sustainability. It operates through information–environment management rather than through individually targeted choice–architecture manipulation [11,12], and whether a specifically designed nudge embedded within this cycle can further improve outcomes is a question for future research. Any such strategy must also attend to the ethical dimension: redirecting tourists toward peripheral areas that lack adequate infrastructure risks displacing overtourism burdens rather than resolving them, and raises questions about the transparency and accountability of content interventions in public information environments [3].

5.5. Limitations and Future Research

This study had several limitations.

Three caveats deserve to be flagged. In terms of measurement, Twitter is only one platform and is much smaller in Thailand than Facebook or LINE [54], with a user base that skews urban, young, and bilingual. Visually oriented platforms such as Instagram and TikTok and messaging services such as LINE and WeChat may generate cascades of different intensities and content compositions; therefore, our estimates should be read as a conservative benchmark for the sub-population that Twitter actually reaches. Moreover, the 2019 data capture a pre-pandemic platform landscape in which Twitter occupied a different competitive position; the subsequent rise of algorithmically curated, video-first platforms such as TikTok may generate cascades of different temporal velocities and content modalities, making our text- and timestamp-based estimates a conservative baseline for the visual-platform era. Twitter sparsity and partial geocoding further qualify the buzz variable, but the EWMA smoothing, district and date fixed effects, and the IV-to-OLS coefficient gap [49] together accommodate this profile.

On the dependent side, the GPS sample is taxi-borne. It over-represents international and upper-middle-class tourists, longer cross-district trips, and evening movement, while under-representing backpackers, public transport commuters, and walking-only tourists. The congestion proxy also captures vehicular rather than pedestrian crowding. Integrating taxi GPS with BTS tap-in/tap-out data, mobile-phone trajectories, or on-site pedestrian surveys would broaden both the sampling frame and the congestion measurement in future work.

On identification, the Bartik instrument is weak in the pre-festival period and we therefore confine IV inference to the During and After windows. The Borusyak–Hull shift-orthogonality test [21] returns a positive correlation that we attribute to the common-shock structure of Songkran; thus, the causal argument rests on the joint strength of the temporal lag, placebo permutation, three-IV convergence, and diffuse Rotemberg weights rather than on any single diagnostic. Because buzz is measured at the district

level rather than the individual level, individual-level causal claims would require an exposure-representativeness assumption that our data do not directly test.

Future work could profitably integrate multiple social media platforms, run experimental designs that manipulate buzz intensity in controlled settings, and follow DMO social media interventions longitudinally to test whether they produce sustained changes in the spatial patterns of tourists. Adding sentiment analysis would also enrich our understanding of how the valence of online discussion differentially shapes spatial choices. Because our buzz variable captures volume rather than valence, we cannot decompose the cascade into attraction-driven and warning-driven components; future work incorporating sentiment classification could test whether negative buzz (e.g., crowding complaints) attenuates or, paradoxically, amplifies the cascade through a curiosity or fear-of-missing-out channel.

6. Conclusions

This study set out to assess whether and under what conditions the digital information environment causally redirects tourists across a city, with the urban festival treated as the empirical setting in which information asymmetry, decision heuristics, and crowd dynamics intersect at heightened intensity. Prior tourism electronic word-of-mouth research had operated almost exclusively at aggregate or correlational resolutions, leaving the individual-level causal pathway and its festival-specific moderation as questions of first-order policy relevance that revealed-preference evidence had not yet addressed.

Drawing on revealed-preference movement traces and geo-located online discussion from the 2019 Bangkok Songkran Festival, the analysis provides evidence that area-level social-media intensity is a causal driver of where tourists actually go, that this influence intensifies under festival-period information asymmetry, and that its potency is conditional on the type of choice. Visibility-driven categories, such as shopping and nightlife, respond strongly to ambient digital chatter, whereas experience-driven categories, such as restaurants and spas, remain anchored to private trust signals that aggregate buzz cannot displace. Once classical measurement error in sparse social-media data is corrected, the magnitude of the cascade proves substantially larger than what conventional correlational benchmarks have reported, suggesting that tourism electronic word-of-mouth evidence has been understating rather than overstating the underlying behavioral response.

Theoretically and methodologically, the contribution is fourfold. The classical cascade tradition is extended from sequential discrete-choice settings into continuous urban spatial choice, showing that herd dynamics operate not only in buy-versus-not-buy decisions but also in where-to-go decisions across geographical alternatives. The focus of electronic word-of-mouth analysis has transitioned from individual venues to urban districts, thereby reconceptualizing the digital information environment as a geographic field rather than as a mere collection of product evaluations. Festival-period information asymmetry and the search-versus-experience distinction, both of which have long been proposed conceptually but never directly tested at this resolution, are established as two empirical boundary conditions under which information cascades materialize most strongly. The Bartik shift-share design, supported by Rotemberg-weight decomposition and the Borusyak–Hull shift-orthogonality diagnostic, is introduced as a workable template for credible causal identification in tourism electronic word-of-mouth research; this methodological contribution complements the substantive findings and can be adopted by future studies in the area without depending on the specific empirical setting analyzed here. Taken together, these four contributions complicate the prevailing twin-transition optimism that digitalization and sustainability advance in tandem, because the same platforms credited with enabling scalable visitor-flow management simultaneously generate organic cascades that reinforce existing spatial concentration; the net contribution to sustainability depends in part on

whether the information environment is actively governed rather than left to amplify market-driven concentration.

For destination management organizations, the findings provide a calibrating set of operational guidance points rather than a turnkey prescription. First, real-time district-level monitoring of online discussion intensity offers a practical early warning instrument for emergent overtourism because behavioral concentration becomes detectable in the digital record before it materializes in the physical record. Second, content-governance interventions should be category-stratified rather than uniform: shopping and nightlife corridors are likely to be the highest-leverage targets for visibility-based redistribution campaigns, while restaurants and spas warrant curated local-experience programming and caution toward any rating-boosting instrument that might signal touristic over-commercialization in categories where authenticity drives selection. Third, festival windows are particularly leveraged temporal triggers for concentrated intervention rather than disruptions to be passively weathered, because tourists become both more responsive to digital signals and more willing to venture beyond their immediate vicinity when overtourism pressure peaks. The after-festival channel remains operative but has a smaller marginal response. Fourth, any information-based redistribution program should be paired with an *ex ante* carrying-capacity assessment of the receiving districts, because redirecting overtourism burdens onto peripheral neighborhoods that lack the infrastructure to absorb them risks converting a sustainability gain into sustainability harm. Whether and how a specifically designed intervention embedded within these guidance points further improves outcomes is a question for prospective intervention research, which the present observational design does not evaluate.

From a sustainability standpoint, the results expose a duality. Organic information cascades concentrate tourist flows in already strained corridors and reinforce localized environmental, social, and economic externalities that work against the Sustainable Development Goals on inclusive urban communities and responsible consumption. Nonetheless, the aggregate spatial distribution moves toward greater equity over the festival cycle, indicating that the same digital information environment, if actively governed, could be leveraged to reverse rather than reinforce concentration dynamics. The natural extension of this work lies in three directions: longitudinal evidence on whether destination-management content interventions translate into sustained spatial redistribution rather than transient response; cross-platform integration that captures the algorithmic-curation regimes of visual-first social media beyond the text-and-timestamp environment examined here; and prospective intervention research that converts the behavioral substrate documented here into evaluated, ethically scoped policy instruments. Sustainable urban tourism in the post-pandemic era will increasingly depend on whether destination managers govern the digital information environment as deliberately as they govern physical infrastructure, and this study provides the behavioral evidence, the measurement architecture, and the operational cycle on which that governance can rest.

Author Contributions: Conceptualization, Y.W. and Z.X.; methodology, Y.W. and Z.X.; software, Y.W.; formal analysis, Y.W. and Z.X.; investigation, Y.W.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Z.X. and Y.W.; visualization, Y.W.; supervision, Z.X.; project administration, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were not required for this study, as it exclusively analyzes secondary, anonymized, and aggregated data from publicly available sources (Twitter API, Google Places API) and a third-party data provider (ITIC Foundation). No direct interaction with human subjects was conducted, and no individual can be identified from the data used. This exemption is consistent with Article 32 of China’s *Measures for Ethical Review of Life Sciences and Medical Research Involving Human Subjects* (2023), U.S. 45 CFR 46.104(d)(4), and EU GDPR Recital 26 regarding anonymized data.

Informed Consent Statement: Not applicable. The study uses only secondary, anonymized data with no direct contact with identifiable individuals.

Data Availability Statement: The data analyzed in this study are openly available under a CC-BY 4.0 license in the Zenodo repository at <https://doi.org/10.5281/zenodo.19152102>. The taxi-fleet records were originally provided by the ITIC Foundation (<https://org.iticfoundation.org/> (accessed on 1 March 2026)); all records are anonymized through identifier hashing, and the Twitter component contains no raw tweet text, profile information, or other content that could enable re-identification.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DMO	Destination Management Organization
eWOM	Electronic Word-of-Mouth
EWMA	Exponentially Weighted Moving Average
GPS	Global Positioning System
HHI	Herfindahl–Hirschman Index
MNL	Multinomial (Conditional) Logit
POI	Point of Interest
RUM	Random Utility Maximization
SDG	Sustainable Development Goal

Appendix A. Supplementary Robustness Results

This appendix collects the supplementary robustness and diagnostic analyses introduced in response to the reviewer comments. Scripts and summary outputs are available in this Appendix A.

Table A1. Buzz coefficient robustness across alternative specifications.

Specification	N Trips	Buzz Coef	z
<i>Panel A: POI-to-district buffer</i>			
B0: baseline (no buffer filter)	95,682	+0.0094 ***	18.10
B1: drop POIs with ≤ 0.5 km ambiguity	68,977	+0.0117 ***	19.41
B1: drop POIs with ≤ 1.0 km ambiguity	44,474	+0.0116 ***	14.85
B1: drop POIs with ≤ 2.0 km ambiguity	18,993	+0.0030 *	2.09
<i>Panel B: Choice-set $K \times P$ grid</i>			
$K = 10, P = 75$	95,692	+0.0098 ***	17.68
$K = 10, P = 85$	95,692	+0.0106 ***	18.62
$K = 10, P = 90$	95,692	+0.0102 ***	17.36
$K = 10, P = 95$	95,692	+0.0109 ***	17.84
$K = 20, P = 75$	95,692	+0.0100 ***	18.68
$K = 20, P = 85$	95,692	+0.0106 ***	19.26
$K = 20, P = 90$ (baseline)	95,692	+0.0097 ***	18.20
$K = 20, P = 95$	95,692	+0.0108 ***	18.73

Table A1. *Cont.*

Specification	N Trips	Buzz Coef	z
$K = 30, P = 90$	95,692	+0.0101 ***	18.26
$K = 30, P = 95$	95,692	+0.0105 ***	18.59

Panel A varies the buffer-distance filter on the nearest-centroid POI-to-district assignment. Panel B varies the number of non-chosen alternatives (K) and the travel-distance percentile threshold (P). *** $p < 0.001$, * $p < 0.05$.

Table A2. Bartik IV cross-specification Rotemberg weights and cross-IV comparison.

Share	1st-Stage F	IV Buzz Coef	Top-1 Weight (District)	Effective N
All-POI-share (primary)	132.03	+0.0226	0.317 (Bang Na)	7.4
Tweet-share (robustness)	124.14	+0.0215	0.661 (Pathum Wan)	1.9
Buzz-POI-share (original)	148.83	+0.0218	0.460 (Khlomg Toei)	3.9

Table A3. POI-category analysis of Rating and Buzz coefficients.

	N Trips	Rating Coef	Buzz Coef	ρ^2
<i>Panel A: Pooled model with and without POI-category fixed effects</i>				
S1: baseline (no category FE)	95,692	−0.060 ***	+0.010 ***	0.106
S2: + POI-category FE (4 dummies)	95,692	+0.029 ***	+0.010 ***	0.198
<i>Panel B: Category-stratified subsamples</i>				
Restaurant (experience)	29,887	+0.017 ***	+0.0040 ***	—
Spa (experience)	21,221	−0.158 ***	+0.0073 ***	—
Shopping Mall (search)	10,234	+0.364 ***	+0.0210 ***	—
NightClub/Bar (search)	31,225	−0.154 ***	+0.0199 ***	—

Panel A shows that adding POI-category fixed effects flips the sign of the pooled Rating coefficient while leaving Buzz essentially unchanged. Panel B shows that within-category Rating preferences are positive for Restaurant and Shopping Mall but negative for Spa and NightClub/Bar, consistent with authenticity-seeking in the latter pair. Buzz sensitivity is higher in the two search-good categories. *** $p < 0.001$.

Table A4. H4 diagnostics at the district–day level.

Test/Outcome	Statistic	p -Value
<i>Panel A: Formal trend tests on daily Gini series ($N = 21$ days)</i>		
Linear trend (slope = $-0.00172/\text{day}$)	$t = -4.48$	0.0003
Mann–Kendall monotonic trend	$z = -3.02$	0.0025
One-way ANOVA (3 periods)	$F = 7.88$	0.0035
Kruskal–Wallis (3 periods)	$H = 9.11$	0.0105
t -test Before vs. After period means	$t = 3.66$	0.0033
Bootstrap 95% CI of Before – After difference	$[+0.013, +0.037]$	0.003
<i>Panel B: Panel 2SLS on district–day outcomes (All-POI-share IV)</i>		
Congestion (with district and date FE)	$\hat{\beta}^{\text{IV}} = -0.0025$	0.155 (null)
$\log(1 + \text{visits})$ (with district and date FE)	$\hat{\beta}^{\text{IV}} = +0.027^{**}$	0.015 (causal)
Visits in levels (with district and date FE)	$\hat{\beta}^{\text{IV}} = +4.67^{**}$	0.002 (causal)

Panel A confirms that the daily Gini coefficient declines significantly over the 21-day window. Panel B shows that Bartik IV identifies a causal visitation channel while leaving within-district vehicular congestion null at the current measurement resolution, consistent with the modest share of tourist trips in total roadway traffic. ** $p < 0.01$.

References

1. Getz, D. The nature and scope of festival studies. *Int. J. Event Manag. Res.* **2010**, *5*, 1–47. [[CrossRef](#)]
2. De Geus, S.; Richards, G.; Toepoel, V. Conceptualisation and operationalisation of event and festival experiences: Creation of an event experience scale. *Scand. J. Hosp. Tour.* **2016**, *16*, 274–296. [[CrossRef](#)]
3. Koens, K.; Postma, A.; Papp, B. Is overtourism overused? Understanding the impact of tourism in a city context. *Sustainability* **2018**, *10*, 4384. [[CrossRef](#)]

4. Capocchi, A.; Vallone, C.; Amaduzzi, A.; Pierotti, M. Is 'overtourism' a new issue in tourism development or just a new term for an already known phenomenon? *Curr. Issues Tour.* **2020**, *23*, 2235–2239. [[CrossRef](#)]
5. Milano, C.; Novelli, M.; Cheer, J.M. Overtourism and degrowth: A social movements perspective. *J. Sustain. Tour.* **2019**, *27*, 1857–1875. [[CrossRef](#)]
6. Thaler, R.H.; Sunstein, C.R. *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Rev. ed.; Penguin Books: London, UK, 2009.
7. Gössling, S. Tourism, technology and ICT: A critical review of affordances and concessions. *J. Sustain. Tour.* **2021**, *29*, 733–750. [[CrossRef](#)]
8. Tiago, F.; Gil, A.; Stemberger, S.; Borges-Tiago, T. Digital sustainability communication in tourism. *J. Innov. Knowl.* **2021**, *6*, 27–34. [[CrossRef](#)]
9. Femenia-Serra, F.; Gretzel, U. Influencer marketing for tourism destinations: Lessons from a mature destination. In *Proceedings of the Information and Communication Technologies in Tourism 2020: Proceedings of the International Conference, Surrey, UK, 8–10 January 2020*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 65–78.
10. Huang, Y.; Qian, L.; Tu, H. When social media exposure backfires on travel: The role of social media-induced travel anxiety. *Tour. Manag.* **2025**, *110*, 105163. [[CrossRef](#)]
11. Jesse, M.; Jannach, D. Digital nudging with recommender systems: Survey and future directions. *Comput. Hum. Behav. Rep.* **2021**, *3*, 100052. [[CrossRef](#)]
12. Tussyadiah, I. A review of research into automation in tourism: Launching the Annals of Tourism Research Curated Collection on Artificial Intelligence and Robotics in Tourism. *Ann. Tour. Res.* **2020**, *81*, 102883. [[CrossRef](#)]
13. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [[CrossRef](#)]
14. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A big data analytics method for tourist behaviour analysis. *Inf. Manag.* **2017**, *54*, 771–785. [[CrossRef](#)]
15. Leung, X.Y.; Sun, J.; Bai, B. Bibliometrics of social media research: A co-citation and co-word analysis. *Int. J. Hosp. Manag.* **2017**, *66*, 35–45. [[CrossRef](#)]
16. Banerjee, A.V. A simple model of herd behavior. *Q. J. Econ.* **1992**, *107*, 797–817. [[CrossRef](#)]
17. Gössling, S.; Scott, D.; Hall, C.M. Pandemics, tourism and global change: A rapid assessment of COVID-19. *J. Sustain. Tour.* **2020**, *29*, 1–20. [[CrossRef](#)]
18. Sigala, M. Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *J. Bus. Res.* **2020**, *117*, 312–321. [[CrossRef](#)]
19. Bikhchandani, S.; Hirshleifer, D.; Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econ.* **1992**, *100*, 992–1026. [[CrossRef](#)]
20. Goldsmith-Pinkham, P.; Sorkin, I.; Swift, H. Bartik instruments: What, when, why, and how. *Am. Econ. Rev.* **2020**, *110*, 2586–2624. [[CrossRef](#)]
21. Borusyak, K.; Hull, P.; Jaravel, X. Quasi-experimental shift-share research designs. *Rev. Econ. Stud.* **2022**, *89*, 181–213. [[CrossRef](#)]
22. Sotiriadis, M.D. Sharing tourism experiences in social media: A literature review and a set of suggested business strategies. *Int. J. Contemp. Hosp. Manag.* **2017**, *29*, 179–225. [[CrossRef](#)]
23. Yang, X.; Lin, Z.; Kargar, M.; Djafarova, E. The echoes of social media friends' travels: Social influence and venue selection in a hyperconnected world. *Humanit. Soc. Sci. Commun.* **2025**, *12*, 1069. [[CrossRef](#)]
24. Song, Q.; Wondirad, A. Demystifying the nexus between social media usage and overtourism: Evidence from Hangzhou, China. *Asia Pac. J. Tour. Res.* **2023**, *28*, 364–385. [[CrossRef](#)]
25. Filieri, R.; Lin, Z.; Pino, G.; Alguezaui, S.; Inversini, A. The role of visual cues in eWOM on consumers' behavioral intention and decisions. *J. Bus. Res.* **2021**, *135*, 663–675. [[CrossRef](#)]
26. Liu, J.; Wang, C.; Zhang, T.C. Exploring social media affordances in tourist destination image formation: A study on China's rural tourism destination. *Tour. Manag.* **2024**, *101*, 104843. [[CrossRef](#)]
27. Raun, J.; Ahas, R.; Tiru, M. Measuring tourism destinations using mobile tracking data. *Tour. Manag.* **2016**, *57*, 202–212. [[CrossRef](#)]
28. Boto-García, D.; Balado-Naves, R.; Suárez-Fernández, S. A behavioral microeconomics model of tourist destination choice under loss aversion. *J. Hosp. Tour. Res.* **2025**, *49*, 1197–1211. [[CrossRef](#)]
29. Hardy, A.; Shoval, N. 25 years of tourist tracking: A geographical perspective. *Tour. Geogr.* **2025**, *27*, 851–862. [[CrossRef](#)]
30. Shoval, N.; Ahas, R. The use of tracking technologies in tourism research: The first decade. *Tour. Geogr.* **2016**, *18*, 587–606. [[CrossRef](#)]
31. Chen, J.; Shoval, N.; Stantic, B. Tracking tourist mobility in the big data era: Insights from data, theory, and future directions. *Tour. Geogr.* **2024**, *26*, 1381–1411. [[CrossRef](#)]
32. Zhou, F.; Xu, X.; Trajcevski, G.; Zhang, K. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Comput. Surv.* **2021**, *54*, 1–36. . [[CrossRef](#)]

33. Shen, C. Social media marketing and digital influence for visitor flow management in sustainable heritage tourism. *Sci. Rep.* **2025**, *15*, 45767. [[CrossRef](#)] [[PubMed](#)]
34. Huang, Q.; Li, H.; Li, M.Y.; Li, X.Y. A Study on the Conforming Behavior of Online Tourism Consumers from the Perspective of Information Cascades. *J. Qual. Assur. Hosp. Tour.* **2025**, 1–35. [[CrossRef](#)]
35. Weinmann, M.; Schneider, C.; vom Brocke, J. Digital nudging. *Bus. Inf. Syst. Eng.* **2016**, *58*, 433–436. [[CrossRef](#)]
36. Juvan, E.; Ring, A.; Leisch, F.; Dolnicar, S. Tourist segments' justifications for behaving in an environmentally unsustainable way. *J. Sustain. Tour.* **2016**, *24*, 1506–1522. [[CrossRef](#)]
37. Dolnicar, S. Designing for more environmentally friendly tourism. *Ann. Tour. Res.* **2020**, *84*, 102933. [[CrossRef](#)]
38. Ni, X.; Wang, D.; Chang, J.; Li, H. Digital nudging for sustainable tourist behavior in new media. *Tour. Manag.* **2025**, *107*, 105087. [[CrossRef](#)]
39. Türkcan, B. Sustainable urban tourism from the perspectives of overtourism and smart tourism: A systematic literature review. *J. Travel Tour. Res.* **2024**, *25*, 112–156.
40. Banerjee, S.; George, A. Identifying overtourism & spill-over tourism using ST-DBSCAN analysis for sustainable management of tourism. *Curr. Issues Tour.* **2025**, *28*, 2927–2947.
41. Mashkov, R.; Shoval, N. Using high-resolution GPS data to create a tourism Intensity-Density Index. *Tour. Geogr.* **2023**, *25*, 1657–1678. [[CrossRef](#)]
42. Nelson, P. Information and consumer behavior. *J. Political Econ.* **1970**, *78*, 311–329. [[CrossRef](#)]
43. Klein, L.R. Evaluating the potential of interactive media through a new lens: Search versus experience goods. *J. Bus. Res.* **1998**, *41*, 195–203. [[CrossRef](#)]
44. Lança, M.; Silva, J.A.; Andraz, J.; Nunes, R.; Pereira, L.N. The moderating role of tourism intensity on residents' intentions towards pro-tourism behaviours. *J. Sustain. Tour.* **2025**, *33*, 822–839. [[CrossRef](#)]
45. McFadden, D. Modelling the Choice of Residential Location. In *Spatial Interaction Theory and Planning Models*; Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J.W., Eds.; North-Holland: Amsterdam, The Netherlands, 1978; pp. 75–96.
46. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
47. Nerella, S.; Bhat, C.R. Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transp. Res. Rec.* **2004**, *1894*, 11–19. [[CrossRef](#)]
48. Petrin, A.; Train, K. A control function approach to endogeneity in consumer choice models. *J. Mark. Res.* **2010**, *47*, 3–13. [[CrossRef](#)]
49. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*; MIT Press: Cambridge, MA, USA, 2010.
50. Staiger, D.O.; Stock, J.H. Instrumental variables regression with weak instruments. *Econometrica* **1997**, *65*, 557–586. [[CrossRef](#)]
51. Sun, Y.Y.; Faturay, F.; Lenzen, M.; Gössling, S.; Higham, J. Drivers of global tourism carbon emissions. *Nat. Commun.* **2024**, *15*, 10384. [[CrossRef](#)]
52. Cavallin Toscani, A.; Vendraminelli, L.; Vinelli, A. Environmental sustainability in the event industry: A systematic review and a research agenda. *J. Sustain. Tour.* **2024**, *32*, 2663–2697. [[CrossRef](#)]
53. Wu, W.; Xu, C.; Zhao, M.; Li, X.; Law, R. Digital tourism and smart development: State-of-the-art review. *Sustainability* **2024**, *16*, 10382. [[CrossRef](#)]
54. Kemp, S. Digital 2019: Thailand. DataReportal. Monthly Active Users for Twitter (~10 M), Facebook (~51 M), LINE (~44 M) in Thailand. 2019. Available online: <https://datareportal.com/reports/digital-2019-thailand> (accessed on 1 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.