International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 9, Issue 3
DOI: https://doi.org/10.15157/IJITIS.2026.9.3.1785-1851
Received: 12.03.2026; Revised: 22.06.2026; Accepted: 24.08.2026

Research Article

Data-Driven Modeling of Tourism Demand in an Emerging Mediterranean Destination: A Google Trends Network, Granger Causality, and Forecasting Framework

Indrit Bahollı[1], Endri Plasari[1]*, Samel Kruja[2], Sidorela Marini[3], Elvin Meka[4]

[1]Department of Computer Engineering, Tirana Business University College, Tirana, Albania
[2]Department of Business Informatics, Tirana Business University College, Tirana, Albania
[3]Center for Economic and Legal Research and Studies, Tirana Business University College, Tirana, Albania
[4]Department of Management and Finance, Tirana Business University College, Tirana, Albania
***endri.plasari@tbu.edu.al**

Abstract

This study proposes an integrated, data-driven systems framework for characterizing and forecasting tourism demand dynamics in Vlora, Albania, an emerging Mediterranean destination. Using 61 monthly observations from February 2021 to February 2026 and 15 tourism-related Google Trends search terms across four thematic attraction dimensions, the framework integrates correlation-network analysis, de-seasonalization, false discovery rate (FDR)-corrected Granger causality, null-model benchmarking, Network Vector Autoregression (Network-VAR), Transfer Entropy, and a newly developed Structural Demand Integration (SDI) Index. The raw network exhibits high connectivity (density = 0.835) and differs significantly from random co-movement ($p = 0.001$). De-seasonalization reduces 91% of network edges and reveals a robust six-node cross-category structural core, while Kanina Castle becomes fully isolated. Following FDR correction, 18 of 210 Granger relationships remain significant, identifying Gjipe Beach, Sazan Island, and the Bay of Grama as principal demand-leading nodes. Transfer Entropy further supports these directional relationships. The SDI Index identifies St. Mary's Monastery, restaurants in Vlora, and Sazan Island as the most strongly integrated tourism assets. Network-VAR forecasting with Granger-leading nodes significantly outperforms ARIMA (Diebold–Mariano $p = 0.0019$). Validation against official INSTAT arrival statistics confirms a strong association with physical tourism demand, including after de-seasonalization ($r = 0.82$). The framework provides a transferable systems-analytics approach for tourism-demand analysis and forecasting in data-scarce Mediterranean destinations.

Keywords: Google Trends; Tourism Network Analysis; Demand Seasonality; Destination Management.

INTRODUCTION

The international travel and tourism industry has traditionally utilized advanced measurement and monitoring methods, such as the official statistic on international visits to destinations and household expenditure surveys, as the benchmarks by which it relies upon for determining its destination demand. Although these two traditional measurement sources continue to provide critical macro-level information and insights into the industry, as the destination management professionals seek to provide their tourism business clients with the most detailed information regarding their businesses' tourist demand for specific locations in a more timely manner, the traditional measurement methods have several key limitations, primarily lag period to establish data, high costs associated with producing the information, and the inability of these sources to support detailed destination management types of analytical models. The emergence of big data and, more specifically, the availability of large volumes of online search activity data, shows great promise as a way to develop new and additional metrics, and, in particular, provides destination management professionals with the ability to gain insights into near-real-time tourist demand [1, 2]. One of the most notable examples of this is the use of Google Trends as a publicly available dataset based on search activity. The Google search database has been validated in multiple international contexts as an accurate proxy for measuring tourism demand. The Google search database has proven to be consistently accurate in forecasting actual incoming tourists, as well as establishing the types of tourist interests associated with specific destinations [3-5].

Albania, and more specifically the Vlora region, offers a case study of uncontrolled, rapid tourism growth in an emerging Mediterranean destination. Tourism in Albania developed substantially throughout the 2010s, due to an extended period of international isolation, and since the pandemic, the country has seen record numbers of tourist arrivals: 6.8 million international tourists in 2022, and 30.6% increase in visitor traffic through the first seven months of 2023 compared to a year earlier. In 2023, tourists spent approximately €4.2 billion (i.e., a 47% increase over 2022), making tourism the largest contributor to Albanian export revenues [6, 7]. As the gateway to the Albanian Riviera and the focus of new infrastructure development (including the Vlora International Airport), the Vlora region is at the heart of this expansion. However, how the structure of tourism demand is organized in the region, how various destination elements relate to and attract complementary and independent demand, and to what extent demand is concentrated in the summer months have yet to be characterized systematically.

The lack of observable physical demand is especially critical due mostly to the seasonal conditions in the region; for most visitors (approximately 75% annually), July and August are the primary months when people travel to the coast of Albania. This means that during these key months, there is significant pressure on destination areas, while the remaining regions experience a higher degree of underutilization and support [6]. It has been generally accepted in prior studies that the extreme seasonality of coastal destinations along the Mediterranean Sea negatively impacts their sustainability efforts by causing either localized infrastructure overload during peak periods or economic stagnation during off-peak periods. Given this, an important aspect of addressing this challenge is identifying the types of components that affect each season separately, as opposed to those seasonal components that affect year-round visitation – this cannot be accomplished by simply using aggregate arrival statistics [8-10].

To address these issues, this study utilizes a multi-perspective analysis framework that incorporates both static and dynamic approaches. First, the researchers perform a descriptive seasonality analysis of 15 specific tourism-related search phrases within the Vlora region over a 5-year period (February 2021 to February 2026; n=61) (every month during the observation period). Second, they create a correlation-based representation of the different tourism components via the creation of a network including each of the tourism components with edges indicating instances when different tourism components experience approximately the same amount of demand, and testing for multi-threshold robustness, performing seasonality-adjusted bootstrap confidence intervals (based upon 1,000 resampled networks), creating and testing through Pearson/partial correlation networks, and rolling window stability (each window – 36-months). Thirdly, the researchers conducted extensive testing of demand through the use of advanced inferential analysis procedures (STL decomposition with formal seasonal & trend strength scoring, Granger causality testing of STL residuals with Benjamini–Hochberg false discovery rate (FDR) correction for each of the 210 directional pairs, ADF/KPSS stationarity tests). Fourthly, they developed an SDI Index—a composite index aggregating four metrics (degree centrality, seasonal strength, betweenness centrality, and net Granger leadership)—which will allow for a repeatable method to diagnose destination demand. Finally, multiple tests of the study's analytical robustness were conducted by assessing both null model comparisons against random networks and split-sample stability analysis (2021–2023 vs. 2024-2026), as well as COVID-recovery sensitivity tests (excluding 2021-2022).

Research Gaps

While there has been an increased number of papers focusing on tourism and Google Trends applications, there continue to exist four significant gaps that motivate this study.

- Gap 1: Structural vs Predictive Focus: The majority of existing Google Trends tourism research focuses on predicting demand rather than understanding the structure of how demand functions for a particular destination's component system. This gap will be directly addressed in this study using a correlation network analysis of 15 sub-destination keywords to create an empirical map of how the components of each destination keyword are synchronously connected, and which components are synchronously disconnected or isolated. Forecasting models will not provide this information.
- Gap 2: Seasonal Confounding in Network Creation: A non-negligible % of Google Trends network studies apply Pearson correlations to their raw time-series data, thereby conflating structural co-movement and common seasonal cycles. It has been shown by [11, 12] that normalizing Google Trends data can also lead to systematic artefacts that will generate spurious correlations. While [13] used correlation

networks to classify Greek regional destinations and [8] built Mediterranean seasonality networks, neither used de-seasonalization prior to constructing networks. We will directly help fill this gap by comparing the raw time-series and the seasonally adjusted time-series networks using partial correlation analyses so as to control for common drivers and rolling window analysis to assess the temporal stability of the structural relationships between sub-destinations or keywords.

- Gap 3: Dynamic Demand Leader Evidence: While static correlation networks provide empirical evidence of structural co-movement, they have not been used to assess which destinations lead others in demand dynamics. This gap will be directly addressed through the integration of Granger causality tests conducted on stationary STL residuals, with false discovery rate (FDR) correction and against a null model, to provide directional evidence of demand leadership at the sub-destination keyword level in an emerging Mediterranean destination, a combination that, to our knowledge, has not previously been assembled in a single pipeline for this class of destination. The contribution here is integrative rather than methodological invention: each individual technique is established in the literature, but their combination — de-seasonalised correlation networks, formal stationarity testing, FDR-corrected Granger leadership, and a composite SDI instrument — has not been applied together to a sub-destination keyword system of this kind.
- Gap 4: Validating Structural Network Predictions: Finally, most prior tourism demand correlation network studies have not sought to determine whether or not the identified structural leaders are, in fact, predictors of destination demand, vis-a-vis univariate baselines, by conducting some form of formal forecasting comparison (e.g., in the context of the structural analysis we conducted, how do the predictions of the structural leaders stack up against the predictions of the target nation)."

Research Hypotheses

This study develops and evaluates a research agenda based upon four separate research gaps and the use of the analytical framework presented above, and proposes five formal hypotheses:

- H1 (Structure Validity): All bootstrap confirmed correlation networks will maintain a structurally cohesive core after being de-seasonalised (statistically determinable from random networks of the same density - permutation test $p < 0.05$). Thus, all co-movement observed is indicative of actual structural integration; i.e., every co-movement observed is reflective of a structural integration that is not dependent upon a shared seasonal cycle.
- H2 (Natural Attraction Leadership): Natural attraction keywords will possess the greatest seasonal strength values (STL $F_s > 0.85$) and will emerge as the net Granger leadership (positive net leadership score after FDR correction) to both cultural heritage and cuisine categories. This supports their position as the primary 'pull-factors' of the destination system.

- H3 (Cultural Heritage Peripherality): Cultural heritage nodes will have much lower centrality than either natural attraction nodes and core nodes of the destination (one-tailed Welch's t-test; $p < 0.05$; Cohen's $d > 0.5$), revealing that Kanina Castle remains structurally isolated across every robustness check, substantive evidence of failure of demand integration for heritage assets.
- H4 (SDI Index Diagnostic Validity): The SDI Index provides a theoretically and empirically validated measure of the degree of centrality, seasonal strength, betweenness centrality, and Granger leadership for each individual component of a destination and can be used as a replicable diagnostic instrument for newly formed destination management organizations.
- H5 (Predictive Superiority of Network-Based Information): The Network-VAR Model that utilizes Granger identified leadership keys as exogenous variables will produce significantly superior within-system forecasts utilizing the Diebold – Mariano Test compared to an ARIMA baseline ($p < 0.05$).

The contribution of this study is an integrated, validated framework rather than a single new method. Its value lies in assembling and cross-validating established techniques into a coherent diagnostic pipeline, and in demonstrating that pipeline on an emerging destination. Within this framing, the study makes four specific contributions:

First, a framework for developing a Google Trends-based de-seasonalised correlation network that separates structural demand integration from common seasonal patterns. Partial correlation analysis and seasonality-adjusted validation methods are used to identify true co-movement; the structural demand map produced illustrates the dominant demand core of Vlora's tourism system (natural attractions cluster), the structural bridges (Karaburun Peninsula and Jale Beach), and the persistently isolated heritage node (Kanina Castle). The results are robust to threshold variation, seasonal adjustments, and the exclusion of COVID period data, and through split-sample replication.

Secondly, create an SDI Index, a formal composite diagnostic that combines structural network position and dynamic causal leadership to operationalise both the Destination Competitiveness framework [14-16] into actual metrics on the network.

Additionally, a single systematic method of developing a reproducible digital diagnosis of the demand for a destination with little data was produced using a methodological framework that combined ADF/KPSS for stationarity tests, Granger causality with FDR, bootstrap for confidence levels, benchmarking of null results, and rolling windows to analyse the stability of the demand model of the destination.

Fourth, the nearly zero modularity coefficient ($Q = 0.017$) of the bootstrap confirmed network with statistical confirmation against the null models indicates a structurally nearly homogeneous demand system with limited cross-category integration, producing specific recommendations for inter-sectoral product development.

This study is structured as follows. The literature review reexamines existing contributions to forecasting and structural analysis of tourism using Google Trends, network-based analysis of tourism destinations, and seasonality in Mediterranean locations. The resulting SOTA Comparative Synthesis Table provides a basis for comparing these contributions. The methodology describes the procedures used for data collection, keyword selection, and analytical methods, including advanced inferential procedures and SDI index formulation. The results present findings from all levels of analysis, including stationarity diagnostics, FDR-corrected Granger causality, null model benchmarking, and robustness validation. The discussion integrates the findings in relation to the five hypotheses and existing SOTA contributions. Finally, the conclusions summarize the main findings and provide recommendations for future research.

LITERATURE REVIEW

Google Trends as a Tourism Demand Proxy

Studies have found that Google Trends can be a useful tool in order to assist in generating short term forecasts for several economic sectors (automobile sales, housing markets and travel patterns). The general idea behind this is that the higher the search volume for any search term related to tourism, the more likely people are to travel. This is the basis behind a great deal of the research papers written on how to forecast tourism [1]. The research presented in [3] finds that using Google Trends web and image search indices provides for a higher level of accuracy when forecasting tourism demand for Vienna, versus autoregressive models based only on historical data. The research presented in [17] uses MIDAS models, which combine Google Trends data together with more conventional time series data to forecast tourism demand for Caribbean destinations. The results were consistently superior to the forecasts produced by SARIMA forecasting models for the same destinations. Also, the research presented in [18, 19] shows that "the incorporation of Google search data" into support vector regression (SVR) models resulted in "substantial improvements" in nowcasting tourist arrivals in Barbados from the U.K. and Canada. Finally, the research presented in [5] backs up the utility of both monthly and high-frequency Google Trends data as predictors of tourist arrivals and overnight stays in Prague in MIDAS models.

According to [20] who conducted a literature review from 2012 to 2019 of research studies that use tourism forecasting methods built from Google Trends data. Authors in [2] published a broader systematic literature review of all uses of Big Data in Tourism and Hospitality forecasting methods. In summary, they document that nearly all of the studies reviewed have documented that the use of Search Query data improves Predictive Accuracy when predicting Tourism Forecasts compared to Univariate Benchmarking.

Authors in [4] identified spurious correlations in Google Trends Data that resulted from the normalization of Google Trends data and proposed several methods for correcting these problems. The study in [21] identified two Types of Bias when employing results from single-language Google Trends to forecast Tourism to multi-lingual visitor markets. The studies demonstrated that when using Google Trends results, the use of Language-

Adjusted and Platform-Corrected data greatly exceeds the predictive accuracy of using unadjusted indicators.

Authors in [11, 12] documented a systematic examination of inconsistency stemming from the Normalization Mechanism of Google Trends and the Sampling Processes of Google Trends, and therefore called into question the reliability of using Google Trends data to measure Economic Activity. Collectively, these studies provide the basis for the application of Seasonally Adjusted, Partial-Correlation Validated, and Rolling Window Stability Analysis in this Study.

A similar stream of literature has advanced beyond forecasting, taking advantage of Google Trends data for structural destination analysis. The study in [22] utilized Google Trend's search data to examining tourist online behavior to enhance their demand forecasting at Åre, Sweden. Their study showed that keyword-level analysis can reveal which components of the destination contribute to interest online. Similarly, [23] made the broader theoretical case for ICT-enabled destination intelligence, reporting that internet-derived data would allow destination management to move beyond reactiveness and toward anticipatory management practices. More recently, studies exploited multi-keyword Google Trends datasets to demonstrate relationships amongst tourism components and identify thematic clusters of tourism demand systems. The present study adopts this approach but builds on this connectedness by introducing formal stationarity testing, de-seasonalization, and partial correlation analysis to more extensively establish that the identified structural relationships represent true co-movement of demand and not just seasonal cycles in common.

The present study adds to this literature by addressing an identified gap noted across all of the reviews focused on forecasting - the lack of structural destination demand mapping at the sub-destination keyword level in emerging Mediterranean and Adriatic systems. The corresponding forecasting studies reviewed all used Google Trends data as leading indicators of aggregate arrivals; here, we use synchronous search behavior at the keyword level as a window into the structural organization of destination demand, and reveal insights into which components are congruently integrated, which are disconnected, and which may provide directional leadership in demand patterns - points that standard forecasting does not relate to.

Network Analysis in Tourism Research

For more than twenty years, there has been a solid presence of network analysis in the tourism field. Some network analysis applications used in tourism research include international tourism flows, as well as the functioning of destination marketing organizations, mobility of tourists, and co-visitation to tourist attractions. Authors in [23] employ social networks analysis to illustrate how international tourism networks have changed over time, as demonstrated by the correlation between betweenness and eigenvector centrality with GDP and population. In the area of tourism, [24] offer a comprehensive review of social network analysis applications and determine that betweenness centrality is a key measure of structural brokerage within the overall network by establishing that nodes that connect two parts of the network that would otherwise not connect play a disproportionate role in the flow of information and integrating the demand.

Through the use of network analysis, degree centrality was used to identify anchor attractions that are likely to collaborate amongst each other, in the study by [25]. The structure and the way in which the attractions are connected in Shanghai demonstrate tourist mobility and policy decisions regarding tourism. In the case of [26], the study identified a hub-periphery tree-type structure, consisting of a high level of clustering, short diameter, and positive correlations among the various measures of centrality, while creating a model to establish the connections of tourists through the various activities in Hong Kong. Through the work of [27] and by understanding how the network governance perspective can lead to the activation of tourist attractions from the periphery as part of the tourism destination development process (the results of the present study, which found cultural heritage on the periphery), these findings support the overall theme of how the centrality of any given node (the anchor attraction) predicts the behaviour of tourists and their influences on local tourism policy, as found by the studies referenced above and is the theoretical basis for determining how the uses of both the betweenness centrality and degree centrality can both be used as indicators of the potential to integrate demand in the present study.

Correlation-based approaches to constructing networks are one of the methodological sub-traditions that are particularly relevant to this study. Pearson Correlation Networks were first used by [13] to classify Greek regional destinations by tourism typology and demonstrate that the topology of the resulting networks was consistent with the geographic and socio-economic characteristics of the destinations. The study in [8] applied this methodology to perform a comparative seasonality analysis between Mediterranean countries using overnight stay time series data to create networks at the country level. This study will extend both contributions by (i) creating networks at the sub-destination keyword level as opposed to regional or national aggregates, (ii) applying de-seasonalization techniques before network construction so as to eliminate the seasonal confound identified by [11, 12], and (iii) validating the network created in this study through the use of bootstrap confidence intervals, null model comparisons, rolling window stability analysis, and partial correlation analysis; none of which were used in the studies conducted by [13].

Predictive validation for network-derived structural claims necessitates using established predictions as baselines against which to compare network-derived forecasts. Google-augmented MIDAS models were found to outperform traditional ARIMA models for forecasting tourist arrivals to Caribbean destinations [17]. Authors in [3] showed that the use of SARIMA-GT hybrid models increased the accuracy of forecasts for Europe's urban attractions. To date, no formal testing has been done with respect to a Network-VAR model using network-identified Granger leaders as exogenous regressors within the context of sub-destination keyword systems. The Diebold-Mariano test [28] provides the framework for making statistical inferences related to forecasting accuracy across differing models, while Transfer Entropy [29] provides a complementary means of measuring non-linear information flow that is not captured by linear measures like Granger Causality.

Tourism Seasonality in Mediterranean and Emerging Coastal Destinations

"Seasonality" refers to the systematic concentration of tourism demand within certain time periods. This characteristic is among the most fundamental and structural challenges faced by Mediterranean Coastal Tourism and is found to be faced by all destination types throughout the Mediterranean region.

[9] Established typologies of the causes of seasonality distinguished natural causes, climate, daylight, weather, from institutional causes, holidays, events, and industry practice. The research work in [30] offered conceptual models that included economic, environmental, and social aspects of the negative externalities of seasonality. In the Mediterranean context, [8] sets out seasonal profiles characterizing Mediterranean countries and argues that extreme summer concentrations are de jure a product of the very structural character of destinations principally dedicated to sun-and-sea tourism. Authors in [31] develop a comparison of Mediterranean islands, suggesting that seasonal beach destination tourism systems can not only be treated as Near-Homogeneous network functions, but that islands in which cultural and natural products are articulated have significantly greater seasonal spreading than in those where beach tourism is the overarchingly most prominent motif. Authors in [32, 33] referring to MED islands, an extreme seasonal concentration of the destination is noted. The study in [10] soundly observes that in the North Mediterranean EU coastal regions, non-linearity characterizes the relations between seasonal congestion and regional tourist performances. At the sub-destination scale, [34] have shown that the tourism segment of cultural tourism proceeds by deconcentrating the territorial footprint and area-based congestion of seaside resorts, successfully absorbing visitors in shoulder and off-peak months. Assumptions for decongesting tourism from seasonality, and indeed, the prescriptions for deconcentration of tourist visitation systems have been described in the product diversification modular literature. The study in [33] conceptually describe the “space of opportunity” for developing tourism in new destinations as a second function of two axes based on tourist product/experience’s geographic location. Disparities in demand-space and product-space positionality give rise to alternative scenarios for tourism development, the authors describe in detail, one of which is escape, which basically argues that asymmetric growth adaptation of tourism over spatial area is subject to factors of referral between nodes controlled by these connected nodes. The product diversification literature further demonstrates that structural change in the tourism industry can directly drive demand growth and season extension [34]. Gastronomic Tourism Experiences were shown to have a distinct, cross-category role within the frameworks developed by [35], as they are built independently and as 'bundles' across coastal, urban, and heritage experiences. Emphasizing the overlooked yet too easily accessible “there” horizon of meals and shops, [36] showed that Albanian tourism arrival systems are essentially coast-bound. Cultural

heritage and natural resources of the mountainous hinterland remain undeveloped and hence, suggested that infrastructure development, etc., should be "properly balanced between the two and regional heritage planning be emphasized, since otherwise, we are in danger of turning the tourism of our country into a tourism of unbalanced success that reduces our country's economy to a bore on the coast and burden in the land." [7], described the Albanian tourism system and regionally articulated "quality of service offered to tourists and accomplishment of their requests" as understood that for a territory whose tourism amenities could be categorized as raw materials that would not change very much, and growth problems "stagnate the operating (and political) machinery. Beyond seasonality characterization, the interplay between seasonality and destination network structure has received limited attention. Authors in [37] argued that destination systems require integrated management of complementary products to achieve competitive differentiation, but empirical evidence on how seasonal demand concentration shapes structural integration at the component level is scarce in emerging Mediterranean contexts.

Exploiting a multi-method structural process tracing framework, this study that attempts to unspool the special demands of DMOs attempting tourist destination out of season, and off-again, a discrepancy not projected there as much as this same surprise often projected all around but, when extreme seasonal concentration in distance-travelled related words like beach (max ratio up to 53.67 for Jale Beach) is complicated with a near-homogeneous locate-to-state correlation network (Q = 0.017), narratively and policy discussion-wise, when these same variables logically-preclude each other as signs, the information is conditioned uniquely for the DMOs attempting to deconcentrate the high season dependency of Mediterranean beach tourism destinations.

Theoretical Framework: Destination Competitiveness and Digital Demand Intelligence

This research is based upon two different theoretical frameworks. First is the Destination Competitiveness Model [14], later revised by [38], which describes tourist destinations as networks of interrelated resources, supportive resources, services available at the destination, and conditions that encourage tourists' demand. The Push-Pull Framework of Tourist Motivation [15, 16] provides the second theoretical basis for this research. This framework distinguishes between two types of motivation to travel, push (internal) and pull (external). A tourist's search behavior on Google Trends reflects a product of Pull Factors. The synchronization of search patterns in relation to keywords shows the empirical basis of the comparative network constructed in this study. The overlapping search patterns between keyword searches suggest that simultaneous use of multiple destination elements stimulates the same pool of potential tourists. On the other hand, the disconnects within components do not result in the stimulation of a pool of potential tourists. Therefore, these components do not generate sufficient visibility to facilitate tourist visits or match the prevailing Push Motivations.

A third theoretical perspective focuses on the data source itself: research on how tourists search for information online shows that online searches mirror a systematic decision-making process that takes place prior to making a trip [39, 40]. Over time, the movement away from traditional sources of information (such as guidebooks, travel agents) to digital sources (i.e., Internet) has changed how tourists assess and select their destinations; search engines have become the primary means by which tourists research their plans [40]. Additionally, Smart Tourism research provides evidence that the temporal and spatial characteristics of the search query data encode motivational patterns that have a meaningful relationship to the actual flows of visitors [39]. As such, in the current study, when we look at the co-movement of search volume across destination keywords, we can interpret that relationship not simply as a statistical correlation, but as the aggregate digital signature created by tourists evaluating several parts of a single destination system at the same time.

When considered together, these frameworks support the categorization of this study as an additional contribution to the developing body of literature about digital demand intelligence within destination management [2, 3, 20]. The reframing of prediction into structural diagnosis provides the main theoretical contribution of this study; this distinguishes it from the vast majority of the Google Trends research concerning tourism. Specifically, this research contributes to the area of tourism analytics by creating a single analytical framework that integrates seasonal decomposition, network topology, bootstrap validation, and Granger causality to identify structural demand relationships in the emergent nature of certain tourism destinations.

A Theoretical Model Linking Destination Competitiveness to Network Metrics

The framework of this study rests on two established bodies of theory, which it operationalizes rather than restates. The first is the Destination Competitiveness model [14], in which a destination's sustained performance depends on how effectively its individual resources are organized into a coherent, mutually supporting system rather than functioning as isolated assets. The second is Push–Pull theory [15, 16], which distinguishes the general motivations that move a tourist toward travel (push) from the specific attributes of a destination that attract that demand and shape its direction (pull). Both frameworks are conventionally applied at the level of the destination as a whole; the contribution here is to give them a measurable, component-level expression through network metrics derived from search behavior.

The link between theory and measurement runs through two constructs. The first is year-round structural integration: a destination component contributes to competitiveness to the extent that it is connected to the rest of the system (its demand co-moves with many other components) and that this connection is not confined to the peak season. In the present framework, this construct is captured by degree centrality weighted by the inverse of seasonal strength, $DC \times (1 - F_s)$: a node with many connections and low seasonality is a year-round integrator, whereas a node with the same connections but high seasonality contributes to integration only during the summer pulse. This is the network-metric translation of Ritchie and Crouch's "supporting factors and resources," which emphasizes that an asset's value lies in its contribution to the functioning whole across time, not in its peak-season visibility alone.

The second construct is dynamic pull and brokerage: a component is a stronger competitive asset when it not only bridges otherwise separate parts of the system but also leads their demand, initiating movements that other components follow. This is captured by betweenness centrality weighted by the absolute net Granger leadership, BC × ǀNLǀ. Betweenness identifies the structural broker the node whose removal would fragment the network [24] while net Granger leadership identifies directional pull in the Push–Pull sense, distinguishing components that initiate demand from those that merely respond. A bridge that also leads is, in competitiveness terms, a more valuable integrating asset than a bridge that is merely structurally positioned.

The Structural Demand Integration Index is the sum of these two theoretically grounded components, $SDI_i = DC_i \times (1 - F_{s,i}) + BC_i \times |NL_i|$. The index is therefore not an arbitrary aggregation of available metrics but a direct operationalization of two competitiveness constructs year-round integration and dynamic pull — each mapped to the network statistic that measures it. Figure 1 summarizes this chain from theory to metric to diagnostic outcome. The empirical SDI rankings are read back through this model in the Discussion: a high-SDI node such as St. Mary's Monastery functions as a year-round integrator, a structural bridge such as restaurant Vlora expresses the gastronomy-as-pull mechanism, and an isolated node such as Kanina Castle represents the failure of integration that the competitiveness framework predicts will depress a destination's overall performance.

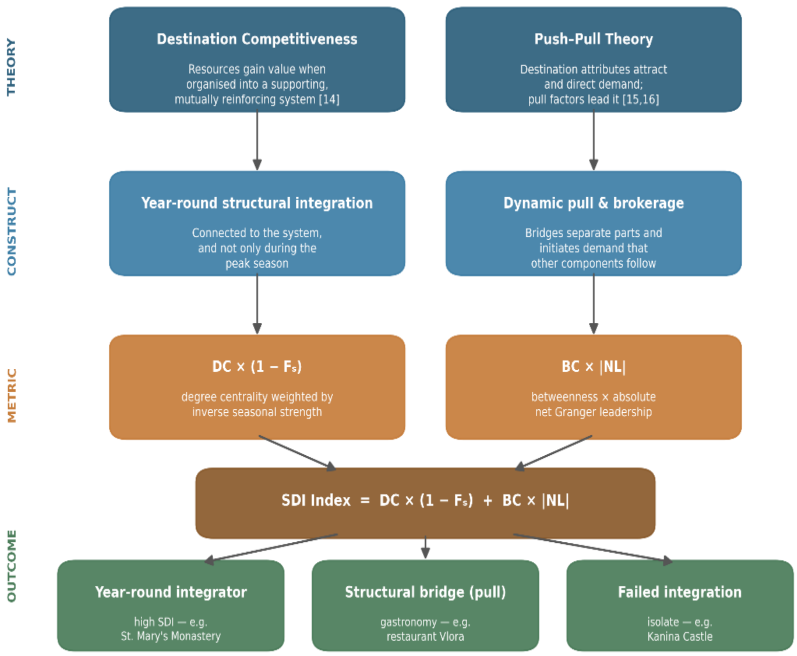


Figure 1. Conceptual Model Linking Destination Competitiveness and Push–Pull Theory to the Network Metrics that Operationalize them and to the Diagnostic Outcomes of the SDI Index.

SOTA Positioning and Research Gaps

The similarities and differences between the studies presented in Table 1 represent the most important information that has been compiled to compare the current project with previous research as it relates to SOTA.

Table 1. Comparative SOTA Synthesis: Positioning the Present Study Within the Literature on Google Trends and Network-Based Tourism Demand Analysis.

Study	Data	Method	Destination Type	De-season.	Null Model	Granger /FDR	SDI/Composite	Gap Addressed
[13]	Official overnight stays	Pearson corr. network	Greek regions (national)	No	No	No	No	Baseline for comparison
[8]	Official overnight stays	Corr. network + seasonality indices	Mediterranean countries	No	No	No	No	Gap 2: no de-season.
[22]	Google Trends	GT forecasting + demand analysis	Åre, Sweden (single dest.)	No	No	No	No	Gap 1: forecasting only
[23]	Official flows	Social network analysis	International flows	No	No	No	No	Gap 3: no keyword level
[25]	Co-visitation data	Network centrality	Shanghai attractions	No	No	No	No	Destination level only
[11, 12]	Google Trends	Statistical consistency analysis	General (no destination)	Yes	No	No	No	Gap 2 identified
Present study	Google Trends (15 keywords)	De-season. corr. network + partial corr. + rolling window + FDR Granger + null models + SDI	Vlora sub-destination (emerging)	Yes	Yes (N=1,000 ER)	Yes (BH-FDR)	Yes (SDI index)	Gaps 1+2+3+4 addressed

Note: De-season. = whether de-seasonalization was applied prior to network construction; BH-FDR = Benjamini–Hochberg false discovery rate correction; SDI = Structural Demand Integration Index.

In comparing the current research with that of previous authors, three major weaknesses can be identified in the existing literature as follows: (i) previous network studies based on Google Trends have only analyzed data at the national or regional level, not at the destination level, (ii) previous studies have not used a time-series approach for network creation or applied the methods of de-seasonalisation, partial correlation and rolling-window stability analysis, and (iii) no previous study has used ADF/KPSS stationarity, FDR-adjusted Granger causality based on STL residuals, null model benchmarks, and a composite SDI, as outlined in the present study. The present study addresses these four gaps by integrating, within one pipeline, the methodological elements that prior studies applied only separately. We do not claim to introduce any of these techniques; the contribution is their combination and mutual validation, and their application to a sub-destination keyword system in an emerging Mediterranean context where such an integrated treatment has been absent.

METHODOLOGY

Study Area and Data Source

The purpose of this research is to explore how tourists` demand for travel products will evolve within the Vlora region, one of Albania's main tourist destinations located along the Adriatic Sea. The region has experienced considerable seasonal variations, rapid increases in tourist visitation, and pressure on its environmental and cultural resources, resulting in an increasing demand for new source regions to create new distributions of travel products. As fine-scale, real-time microdata on tourist arrivals are lacking from the Vlora region, we used a big data approach based on online search patterns. In order to collect sufficient data on travel related online searches we used Google Trends search behavior, which provided us with normalized index values on online searches from a given timeframe. The trends data for this study were obtained using Google Trends from 02.2021 to 02.2026 and included n=61 from a specified area in Albania for the geographic area of interest. Searches performed were limited to Travel category filters to reduce queries that are not related to travel for this study. Data was collected during March 2026, and the specific searches and download parameters used can be found in a supplemental Colab notebook located at a GitHub repository that will be publicly available once this manuscript has been accepted. The files that were exported contained numerous duplicate rows resulting from batch-query normalization therefore, duplicate rows were resolved by taking column-wise maximum values for each date, a common approach for multi-batch Google Trends datasets [4, 21]. The terms restaurant Vlora and restaurants vlore were included as separate nodes in our model due to their respective Pearson correlations ($r = 0.719$, $p < 0.001$) not exceeding the threshold for inclusion in a graph network, indicating that there were two distinct sub-markets identified.

Keyword Selection and Thematic Grouping

The key tourism destinations are the basis for measuring the overall demand for all of the Vlora Riviera's tourism assets, see Table 2. The natural assets that make up the nature-based tourism supply of the area consist of the coastal, marine, and protected areas. The cultural heritage keywords cover all religious and historical tourism attractions located within the region. The demand for experiential and food-related tourism is measured through the use of the cuisine and gastronomy keywords.

These clusters of data enable the examination of the way(s) that the various tourism subsystems interact with each other, rather than simply treating the total demand for a region as one homogeneous entity.

Table 2. Keywords Included in the Analysis by Thematic Category.

#	Keyword	Category	Rationale
1	Vlore	Core Destination	Regional hub and main urban center
2	Palase	Core Destination	Coastal village, southern Riviera
3	Orikum	Core Destination	Southern resort town
4	Dhermi	Core Destination	Major beach destination
5	Himare	Core Destination	Southern Riviera hub
6	Sazan Island	Core Destination	Main island destination of the region
7	Llogara National Park	Natural Attraction	Protected mountain–coastal park
8	Karaburun Peninsula	Natural Attraction	Marine protected area
9	Bay of Grama	Natural Attraction	Scenic coastal heritage site
10	Gjipe Beach	Natural Attraction	Canyon Beach: Adventure Tourism Hub
11	Jale Beach	Natural Attraction	High-traffic seasonal beach
12	St. Mary's Monastery	Cultural Heritage	Zvërnec island monastery (religious tourism)
13	Kanina Castle	Cultural Heritage	Medieval fortress; historical landmark
14	restaurant Vlora	Cuisine & Gastronomy	Gastronomy proxy; dining demand
15	restaurants vlore	Cuisine & Gastronomy	Variant query (retained; $r = 0.72$ with above)

Note: Two cuisine keywords retained as separate nodes ($r = 0.719$, $p < 0.001$).

Seasonal Trend and Descriptive Analysis

The Google Trends search interest for keyword i at time t is represented as x_{it}, where t is a monthly observation ($t = 1,\ldots,61$). The descriptive statistics for each of the time series were computed to describe the intensity and variability of Tourism Interest. Four additional measures were developed to describe the seasonality.

The mean search interest (μ_i) for each of the time series was calculated as the simple arithmetic mean of all of the monthly observations over the 61-month period. The Coefficient of Variation ($CV_i = \sigma_i/\mu_i$) provides a summary statistic about relative variability and provides an absolute measure of how much variability occurred within each tourism interest time series, relative to its mean. The month with the highest mean search interest across the entire study period is referred to as the Peak Month. A Summer Ratio, which is defined as the mean search interest from each June to September, divided by the mean value of all remaining months, has been used to quantify how strong seasonal demand risk is concentrated in the Peak Month. A higher Summer Ratio indicates a greater concentration of demand risk during this peak season.

Correlation-Based Tourism Network Analysis

We performed a correlation-based network analysis to assess structural relationships between tourism components, similarly to approaches outlined by [8, 13]. In creating a network with nodes representing each tourism keyword and edges connecting each pair of nodes, the pairwise Pearson correlation coefficients between time series for each keyword were calculated, and an edge was added between nodes i and j when the magnitude of the Pearson correlation coefficient ($|r_{ij}|$) was greater than or equal to θ, where θ is the chosen correlation threshold. The reason for choosing the Pearson correlation coefficient is based on its historical use in tourism network analyses to provide an understandable measure of the extent to which demand for different tourism products correlates or moves together and is linear. The threshold of θ = 0.70 is also used as the principal threshold for analysis and corresponds with results from prior research conducted on Mediterranean seasonal trends [8, 13]. Additionally, it has been performed in the next sections a sensitivity analysis (robustness check) to assess the robustness of our results to different parameters: θ ∈ {0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.90} for both unadjusted and adjusted data. We also assessed our observed topology against a random assembly of 1,000 networks that were equal in density. A second and separate robustness check was performed by using the precision matrix to produce a partial correlation network (PCN) incorporating common drivers.

To evaluate the robustness of the primary network constructed at θ=0.70, the network was evaluated against other values of θ={0.50,0.60,0.65,0.70,0.75,0.80,0.90}. The results of these analyses indicate that for values of θ from 0.65 to 0.80, the results were generally stable, thus confirming that the most important structural characteristics of the primary network were not artefacts of threshold selection; i.e. a changing threshold did not influence the centrality measures across certain values.

In addition to confirming that the primary network remained stable across θ ∈ {0.65,0.70,0.75,0.80}, three centrality measures were calculated for all nodes in the final network. The degree centrality (DC) measure for node i was calculated as the number of nodes to which node i is connected; the betweenness centrality (BC) measure for node i estimated the extent to which node i lies on any of the shortest paths between two other nodes [40]. The clustering coefficient (CC) measure for node i indicates the number of connections among node i's immediate neighbors and, therefore, the local cohesiveness of those connections.

Seasonality Adjustment and Structural Validation

The tourism industry tends to experience seasonal patterns related to increased interest in certain destinations each summer. When determining if any correlation among the various keyword searches was due to these seasonal influences or if they were due to structural co-movement between keywords, a validation process was applied based on seasonality-adjusted data. To create a seasonality-adjusted series for keyword 'i', the following formula was developed:

$$\tilde{\sigma}_{it} = x_{it} - \mu_i^M \quad (1)$$

where μ_i^M represents the mean level of search interest for keyword 'i' during month 'm(t)' for all available years. By doing this step, the researcher is able to remove the influence of systematic month-to-month fluctuations and determine what portion of the search interest was above or below typical levels of search interest for that specific month of the year. By removing the common seasonal patterns, the remaining correlations between search interest demonstrate co-movement among keywords in addition to the shared peak months of search interest (i.e., summer) [4, 11]. Correlation matrices and network structures were then developed using the adjusted series, and differences between the properties of the adjusted networks and the properties of the raw networks were analysed.

Robustness Analysis and Threshold Sensitivity

A threshold sensitivity analysis using both raw and seasonally-adjusted data was performed to test the stability of the identified network structure. Using several threshold levels to assess the stability of these nodes, the network properties evaluated at each threshold included the number of active nodes, the number of edges, the average degree of nodes, the network density, the number of connected components, and the size of the largest connected component. The confirmation of stability is evident through the persistent consistency of results across a variety of threshold values, and the established patterns of clustering represent tight linkages throughout the Regional Tourism System, rather than being the results of arbitrary parameter selections.

Methodological Limitations

Google Trends data shows how popular a keyword was in the location searched during the time frame chosen, but does not give an accurate representation of the number of visitors coming into a country or how much money was spent on tourism by those visitors. The data is captured only for Albania as a singular location, meaning that it will be impossible to determine if the searches came from people traveling domestically or internationally, or isolate searches for a specific sub-region. Additionally, the amount of exposure via media, marketing, or other events that do not relate directly to travel can also affect the number of times the keyword is searched. As far as the Cultural Heritage system, there are only two keywords, limiting any sort of statistical foundation used to create inferences regarding that System. The external validation relied on the national monthly arrivals as a surrogate measure of the monthly number of visitors to Vlora, as the national Institute of Statistics (INSTAT) did not provide any prefectural monthly data from Vlora. Although a strong correlation can be established at the national level, there is no evidence to support that a correlation exists at a more localized level (i.e., between the two regions). Obtaining a time series of monthly arrivals for Vlora will provide clearer evidence of whether there is indeed a link between Vlora and the Albanian national arrival series, thus creating an opportunity for future studies. Despite these limitations, the use of combined seasonal analyses, network-based infrastructure centrality measures, and multi-threshold robustness provides a validation of a functional and reliable approach to examine the tourism patterns of those geographic regions that do not produce data.

Computational Environment and Analytical Tools

In Google Collaboratory, using Python 3.10, we completed data preprocessing and analysis. We accessed the Google Trends Data manually through the Google Trends interface, selecting Travel as my Category, Albania as my Country, and a Monthly Resolution for each data point. To manipulate the data, we used Pandas and NumPy, then used SciPy for correlation analyses; NetworkX enabled Network Creation, Centrality Calculations, and Robustness Tests; We performed STL Decomposition on the data with *statsmodels.tsa.seasonal.STL*; Granger & Stationarity Tests used *statsmodels.tsa.stattools*; Multiple Testing Adjustments using the number of Tests approach to account for the statistical dependence of all tests were made using *statsmodels.stats.multitest* (Benjamini-Hochberg); Community Detection Methods employed *NetworkX.algorithms.community* (Greedy Modularity & Louvain). To ensure full reproducibility of the entire analytical workflow, including the specific Google Trends query strings, raw CSV files, & all codes with random seed inputs, after acceptance, all outputs will be available via Google Colab Notebook & through the GitHub Repo.

Stationarity Testing and Pre-Condition Verification

Before performing Granger causality tests, all of the keyword series were formally tested for stationarity using two different tests. First, a unit root (non-stationarity) test was performed by means of an Augmented Dickey-Fuller Test (ADF), and second, a stationarity test was performed with the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The null hypothesis of the ADF is that series are not stationary; rejection at an α level of 0.05 indicates that the series is stationary. The null hypothesis for the KPSS test is stationarity; therefore, rejection at the $\alpha = 0.05$ level indicates that the series is non-stationary. By using both of these tests together, a more robust determination of whether the series is stationary can be made than by either test alone; a series is determined to be 'stationary' if the ADF rejects the null hypothesis ($p < 0.05$) and fails to reject the null hypothesis of the KPSS ($p > 0.05$) and a series will be given a 'mixed' evaluation if there is a disagreement between the two tests used. The lag selection for the ADF was done using the Akaike Information Criterion (AIC), and the bandwidth for the KPSS was determined automatically. Because the raw seasonal time series typically do not pass the tests for stationarity due to a strong periodic component, Granger causality tests were performed on the residuals obtained from the STL decomposition instead of the raw time series, as the residual component is made to be stationary, at least by construction, after both the trend and seasonal components have been removed [41].

Partial Correlation Network and Common-Driver Control

A partial correlation network was used to validate Pearson correlations so as to mitigate any risk that seasonality or trending drivers would artificially inflate the correlation coefficients calculated by comparing two keywords with Pearson's coefficient. By utilizing the precision matrix (inverse of the Pearson correlation matrix), the partial correlations were calculated according to the standard formula. The formula for calculating partial correlation is as follows:

$$p_{ij} = -\frac{C_{ij}}{\sqrt{C_{ii} \times C_{jj}}} \tag{2}$$

where *C* is precision matrix. Partial correlation reflects the degree of linear relationship between two keywords after accounting for the effects of all remaining keywords concurrently. The construction of the partial correlation network used a cut-off of $|(p_{ij})| = 0.30$, which represents a lower threshold than that for the Pearson Network and is thus expected to under-represent the number of pairwise structural associations. The comparison between the Pearson and partial correlation networks will indicate which edges are relatively stable in the presence of common-driven confounding compared to those that appear to reflect indirect associations through a shared seasonal pulse.

Null Model Comparison and Network Significance Testing

In order to evaluate whether the characteristics of the correlation network are statistically different from those that would be expected from chance, we constructed two null models for comparison: First, we generated N = 1000 Erdős – Rényi Random Graphs that contained the same number of nodes and edge density as the original correlation network. Second, we created N = 500 Degree-Preserving Rewired Networks (DPRNs) via Double-Edge Swapping, which preserves the degree sequence, but randomizes the connections among nodes. Then we calculated the Clustering Coefficient, the Size of the Largest Connected Component, and the number of Nodes with Positive Betweenness Centrality for each null network. Finally, we calculated the empirical p-value for each null model as a ratio of the number of null networks with one or more values greater than or equal to the metric value that we observed. Thus, we statistically tested the following hypothesis (H1): The correlation network has a significant amount of non-random co-movement when compared to other random correlation networks that have similar density values.

Structural Demand Integration (SDI) Index

To rank structural position, seasonal behaviour and causal dynamic leadership on a single perceived metric, we created the Structural Demand Integration (SDI) Index, which is defined operationally by the theoretical construct of "integration through pull factors" from Destination Competitiveness theory [14]. The SDI Index is defined as follows:

$$SDI_i = DC_i \times (1 - F_{si}) + BC_i \times |\, NL_i \,| \tag{3}$$

where *DC* is the degree centrality (the proportion of connected nodes), *F* is the STL seasonal strength index (0-1), *BC* is the betweenness centrality (normalized), and *NL* is the net Granger leadership score (out-degree - in-degree following FDR correction). The first component of the SDI Index, *DC* × (1 - *F*), encompasses the structural integration that is achieved throughout the year: the node with the greatest number of connections and a low

F value is more integral to the sustainable integration of the destination than a node with the same number of connections, but only contributes to a structural peak season through a high F value. The second component, $BC \times |NL|$, incorporates the value of dynamic brokerage: a structural bridge that causes other nodes to change demand is a greater asset than a bridge that simply connects two points in the system. The theoretical justification for this functional form — why year-round integration is operationalized as $DC \times (1 - F_s)$ and dynamic pull as $BC \times |NL|$ — is summarised in Figure 1. The two components correspond to the 'supporting resources' and 'pull-factor' constructs of the Destination Competitiveness and Push–Pull frameworks, respectively, rather than being chosen for empirical convenience. The SDI Index is a 0-100 normalized value, where 100 indicates the node with the greatest structural integration and dynamic influence on the system. The SDI Index is designed as a diagnostic tool for Destination Marketing Organizations (DMOs) rather than as a forecasting tool, and its validity is based on the statistical validity of the individual measures that make up the SDI Index.

Dynamic Network Stability: Rolling-Window and COVID Sensitivity Analysis

The temporal stability of the identified network structure is evaluated using two additional robustness tests. The rolling window analysis computes correlation networks by using 36-month sliding time windows and looks for changes in the overall density of the network and node-level betweenness centrality through each window to see if the network structure identified across time (61 months) remains constant or if it is only present in one particular part of the time series. The COVID recovery sensitivity analysis also runs an analysis excluding the 2021-2022 (Months 1-24) period since this period overlaps with Albania's recovery from the effects of the COVID-19 pandemic on tourism. In addition to this, the analysis will look at how similar the two network structures are by using the Pearson correlation to compute a correlation value (r) between the degree centrality vectors of both networks. A correlation value of $r > 0.90$ will provide a strong indication of similarity in the two networks' structural properties. A split-sample analysis will also be performed comparing the networks obtained from the 2021-2023 and 2024-2026 sub-periods of the data.

Predictive Validation Framework

The H5 (Predictive Superiority) hypothesis was tested formally by estimating and comparing three competing forecasting models against a held-out test set. The entire sample of sixty-one monthly observations was split up into two windows: a training window covering February 2021 to February 2025 (n=49) and a forecasting window covering March 2025 through February 2026 (n=12).

Model 1 - ARIMA Baseline: Each of the fifteen-keyword series was analysed independently using univariate SARIMA(1,1,1)(1,1,1)12 models. As the Mediterranean tourism cycle has a very distinct annual seasonality (i.e., s=12), this seasonal periodicity was accommodated within the models.

Model 2 - VAR Multivariate: A vector autoregression was fit on the six structurally most important keywords found in the seasonality-adjusted network (Vlore, Dhermi, Palase, Bay of Grama, Himare, restaurant Vlora). The lag order for this model was selected based on the Akaike Information Criterion (AIC); the series were differenced for stationarity before forecasting and then restored to their original levels.

Model 3 - Network-VAR: A SARIMAX(1,1,1)(1,1,1)12 specification was fit for each target keyword contained within the six nodes at the center of the causal structure. The lagged exogenous regressors were the three Granger leaders (Gjipe beach, Sazan Island, Bay of Grama) shown to be statistically significant were included in the model with a lag of one month. This model represents the transformation of the structural causal information obtained from the correlation network into predictive models.

We assessed how accurate our forecasts were according to three categories of errors: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The Diebold-Mariano test was used to perform a pairwise comparison of the forecast error between models and make asymptotic normal inferences about the equality of forecast errors across the competing models, with the estimation method as described in [28].

To quantify nonlinear relationships, we performed Transfer Entropy (TE) calculations on seven of the eighteen FDR-significant Granger pairs. TE is a non-parametric measure of the flow of causal information in both nonlinear and linear causal relationships and was calculated using histogram binning techniques (6 bins, 1 month lag) following the procedures defined in [29]. TE values are reported in bits (log2); TE values that are greater than 0.05 bits signify a significant amount of causal information transfer, and TE values greater than 0.15 bits indicate a very strong causal relationship.

External Validation Against Official Statistics

To determine if the Google Trends data can genuinely measure travel activity as indicated by the physical flow of tourists, we verify the search index against actual arrivals as recorded by the Institute of Statistics of Albania (INSTAT). Vlora prefecture does not generally publish monthly figures for arrivals; therefore, we employed the national monthly figure used for this purpose. Although the mismatch of location is an inherent limitation of this methodology, we feel that it is still valid. Additionally, we established an aggregate search index (here denoted as ASI) that combines three anchors (vlorë, Dhërmi, Palasë) into one series for each month within the time frame of the study (February 2021-February 2026), thus producing an aggregate search index. We performed three separate tests for all indices. In the first test, we computed both the contemporaneous Pearson and Spearman correlation coefficients of the raw and de-seasonalised components (using the monthly calendar adjustment procedure described in equation (1) for de-seasonalising) to remove any correlation due to the common summer cycles from any correlation that occurs purely as a result of true month-to-month movements. In the second test, we used a distributed lag model to test the effect of contemporaneous search index on national monthly arrivals, with a maximum of three-monthly lags. In the third test, we examined

Granger precedence between the search index and national arrivals on de-seasonalised, first-differenced series, after confirming stationarity with the Augmented Dickey–Fuller test. Non-resident arrivals and a fifteen-keyword aggregate search index were used as robustness checks.

RESULTS

Seasonal Dynamics of Tourism-Related Search Interest

According to an analysis of Google Trends data from February 2021 through February 2026 (n = 61), there is a significant amount of annual variation in internet searches made by people interested in visiting the Vlora region of Albania for tourism purposes. There is an exceptionally high frequency during the summer months across all types of searches relating to tourism in the Vlora region. The pattern of each of the 15 search terms and associated search associations is provided in Figure 2 for the four general categories (i.e., beaches/ coastal areas, Llogara National Park, Cultural Heritage & Historical Sites). Peak periods of interest are July-August for the majority of beach and coastal searches made by potential tourists; Llogara National Park search patterns exhibit more 'intermediate' or very gradual intensity over time; Cultural Heritage & Historical Site search patterns fluctuate based primarily on special events most often linked to cultural heritage (e.g., festivals).

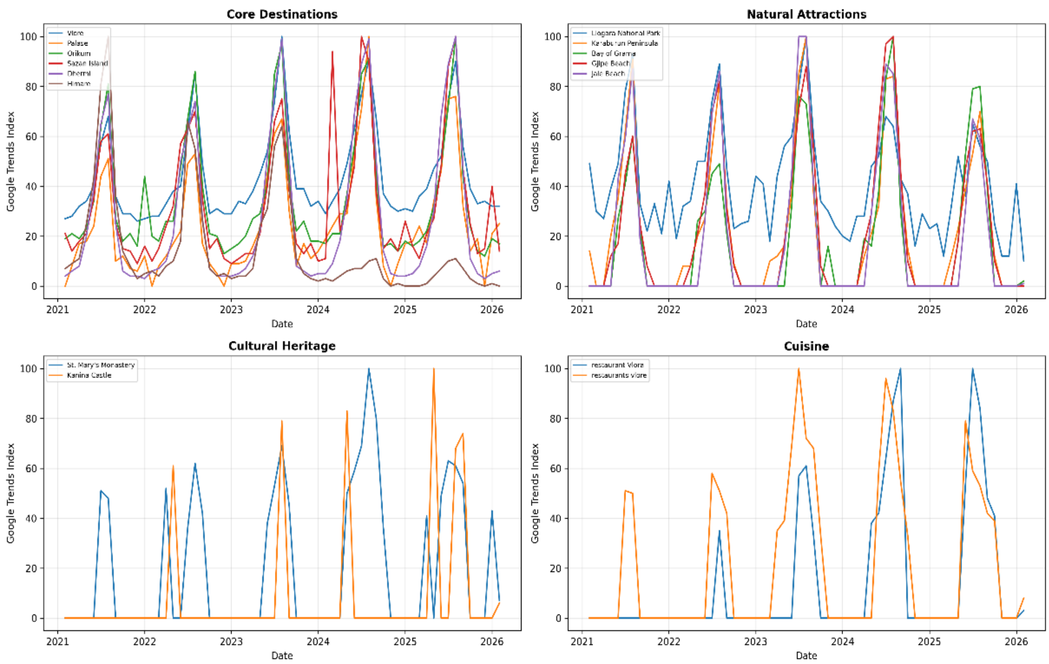


Figure 2. Seasonal Dynamics of Tourism-Related Digital Interest in the Vlora Region (February 2021–February 2026).

Descriptive statistics summarising the seasonal behaviour of all keywords analysed are given in Table 3. Of all series analysed Jale Beach displays the greatest extreme concentration, summer ratio = 53.67 suggesting mean digital interest from June–September is 53 times greater than mean digital interest from the remaining eight months. Bay of Grama (summer ratio = 21.60), restaurant Vlora (19.13), and Gjipe beach (14.14) also display high summer concentration. Llogara National Park by contrast, has the highest mean search interest of all 15 keywords (mean = 40.6) and the lowest Coefficient of Variation (CV = 0.516), with a summer ratio of only 2.08, effectively making it the only empirically documented year-round demand anchor in the system.

Table 3. Seasonal Characteristics of Tourism-Related Search Interest in the Vlora Region (February 2021–February 2026).

Keyword	Mean	SD	CV	Peak Month	Summer Mean	Off-Season Mean	Summer Ratio
NATURE__Jale Beach	19.61	30.94	1.578	8	57.6	1.07	53.673
NATURE__Bay of Grama	18.34	27.02	1.473	8	51.1	2.37	21.599
CUISINE__restaurant Vlora	13.89	27.8	2.002	8	38.25	2	19.125
NATURE__Gjipe Beach	21.49	28.93	1.346	8	57.25	4.05	14.14
CUISINE__restaurants vlore	20.89	30.06	1.439	7	54.35	4.56	11.916
NATURE__Karaburun Peninsula	21.66	28.15	1.3	8	55.4	5.2	10.664
CULTURE__St. Mary's Monastery	19.8	27.88	1.408	8	48.95	5.59	8.764
CORE__Dhermi	27.52	30.29	1.1	8	65.5	9	7.278
CORE__Himare	14.82	21.42	1.445	8	34.3	5.32	6.451
CORE__Palase	24.85	21.86	0.879	8	48.05	13.54	3.55
CORE__Sazan Island	32.38	25.29	0.781	8	59.1	19.34	3.056
CORE__Orikum	34.46	24.18	0.702	8	62.5	20.78	3.008
NATURE__Llogara National Park	40.61	20.95	0.516	8	62.35	30	2.078
CORE__Vlore	43.8	18.48	0.422	8	64.9	33.51	1.937
CULTURE__Kanina Castle	7.72	23.58	3.053	5	11.05	6.1	1.812

Note: Summer ratio = mean search interest June–September / mean search interest remaining months. Keywords sorted by summer ratio (descending).

All types of Natural Attractions have very high variability (CV) values and seasonal (Summer) peaks. High seasonality is also shown in Gjipe Beach (CV = 1.346; Summer Ratio = 14.14) and the Karaburun Peninsula (CV = 1.300; Summer Ratio = 10.66). In contrast, Sazan Island (CV = 0.781; Summer Ratio = 3.06), while also being highly seasonal, has a more stable profile than the previously mentioned beaches, due to a secondary peak occurring during Spring months (indicating marine tourism activities); however, marine tourism is typically unassociated with beach activities.

Core destinations have highly variable seasonal patterns. Dhermi (CV = 1.100; Summer Ratio = 7.28) and Himare (CV = 1.445; Summer Ratio = 6.45) both have extensive fluctuations, with a large summer peak in interest; whereas Vlore has the lowest CV of all core destinations (CV = 0.422), but also the highest average for the category of keywords (Mean = 43.8), indicating that Vlore functions as an urban center and, therefore, draws sporadically throughout the year for demand.

Cultural Heritage Sites have low average search interest and distinct seasonal interest profiles. St. Mary's Monastery receives its highest interest peak in August (Summer Ratio = 8.76) and maintains moderate interest throughout the year. In contrast, Kanina Castle has the lowest average of all keywords and has been shown to have the most erratic pattern of demand (CV = 3.053), peaking in May. This suggests that Kanina Castle is likely to draw only sporadic, not continuous, tourism, unlike St. Mary's Monastic Site or Gjipe Beach. Therefore, the findings confirm Hypothesis 3 that Heritage Nodes will have much less Structural Integration compared to other types of attractions.

Correlation-Based Structure of the Tourism System

Using the Pairwise Pearson correlation methods, 15 of the available words were used to determine the relations between them using a correlation-based network. To do this, we have set a minimum correlation value to determine whether two words should be connected in the correlation-based network.

The value of correlation for the word pairs was set to $|r| \geq 0.70$. All words with a correlation of more than 0.70 are connected in the same network. There is a total of 14 nodes in the correlation-based network, and 76 edges, giving it a density of 0.835 and only one connected component (i.e., all of the nodes are connected in the same component). The only keyword that does not have a connection in the network is Kanina Castle (degree centrality = 0), which has no values greater than 0.70 that correlate with any of the other keywords. A graphic representation of the network can be viewed in Figure 3.

As seen from Table 4, the strongest pair correlations were identified. Dhermi and the Bay of Grama had the highest correlation (r = 0.968), followed by Dhermi and Gjipe Beach (r = 0.966), and Karaburun Peninsula and Jale Beach had a close correlation of 0.963 (r = 0.963).

The close to 1 relationship of Dhermi and the Bay of Grama (Dhermi) with its surrounding areas indicates that these locations have strong synchronized seasonal demand dynamics, as expected in Hypothesis 2.

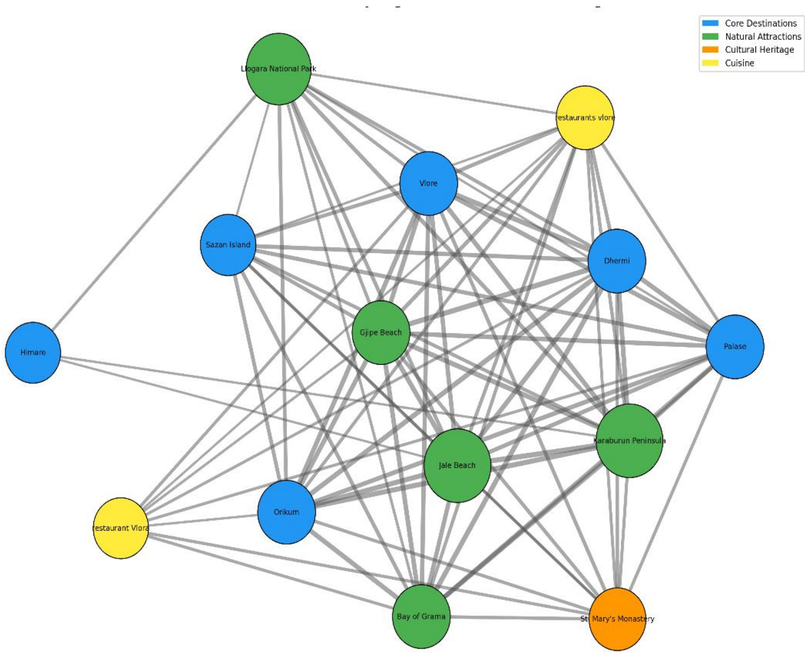


Figure 3. Correlation-Based Tourism Network for the Vlora Region (|R| ≥ 0.70, Raw Data). Node Size is Proportional to Betweenness Centrality. Node Colour Indicates Thematic Category (See Legend). Edge Thickness Reflects Correlation Strength.

Table 4. Strongest Pairwise Correlations Between Tourism Keyword Series (|R| ≥ 0.70, Raw Data, Top 20 Pairs).

Keyword_1	**Keyword_2**	**Pearson r**
CORE__Dhermi	NATURE__Bay of Grama	0.9679
CORE__Dhermi	NATURE__Gjipe Beach	0.966
NATURE__Karaburun Peninsula	NATURE__Jale Beach	0.9631
CORE__Orikum	CORE__Dhermi	0.9571
CORE__Vlore	CORE__Dhermi	0.9548
CORE__Orikum	NATURE__Karaburun Peninsula	0.9538
NATURE__Gjipe Beach	NATURE__Jale Beach	0.9523
CORE__Dhermi	NATURE__Karaburun Peninsula	0.9516
CORE__Orikum	NATURE__Jale Beach	0.9487
NATURE__Bay of Grama	NATURE__Gjipe Beach	0.9484
CORE__Vlore	NATURE__Gjipe Beach	0.948
CORE__Dhermi	NATURE__Jale Beach	0.9465
NATURE__Karaburun Peninsula	NATURE__Gjipe Beach	0.9459
CORE__Vlore	CORE__Orikum	0.9453
CORE__Orikum	NATURE__Bay of Grama	0.9386
CORE__Orikum	NATURE__Gjipe Beach	0.9385
CORE__Vlore	NATURE__Karaburun Peninsula	0.9374
CORE__Palase	CORE__Dhermi	0.9348
CORE__Palase	NATURE__Bay of Grama	0.9321
NATURE__Karaburun Peninsula	NATURE__Bay of Grama	0.9289

Table 5 shows the node centrality metrics. These metrics illustrate how the keywords can be organized in a linear hierarchy. Nine of the keywords have the highest degree centrality: Karaburun Peninsula, Gjipe Beach, Vlore, Dhermi, Palase, Orikum, Bay of Grama, Jale Beach, and restaurants vlore, all having a *DC* value of 0.923. This indicates that these 9 nodes are connected to 12 out of the remaining 13 active nodes. In the second tier, Sazan Island, Llogara National Park, and St. Mary's Monastery have a *DC* value of 0.846. Restaurant Vlora has a DC value of 0.615 and Himare a *DC* value of 0.231, thus they are positioned on the periphery of the core structural network. Betweenness centrality is low, with only two nodes having BC values of over 0.046, and those are Karaburun Peninsula and Jale Beach. The next is Llogara National Park with a BC of 0.038. Kanina Castle has a *DC* value and a *BC* value of 0. It is therefore completely isolated for structural purposes. Therefore, there is an almost uniform distribution of nodes with high degree centrality as demonstrated by the existence of an extremely dense graph (graph density = 0.835), most node pairs are directly connected; very few node pairs need to use intermediate nodes for routing, and this was further confirmed by the null model analysis.

Table 5. Node-Level Centrality Metrics for the Tourism Correlation Network (|R| ≥ 0.70, Raw Data). Nodes Sorted by Degree Centrality.

Keyword	Degree Centrality	Betweenness Centrality	Clustering Coefficient
NATURE__Karaburun Peninsula	0.9231	0.0463	0.7748
NATURE__Gjipe Beach	0.9231	0.0096	0.8424
CORE__Vlore	0.9231	0.0096	0.8407
CORE__Dhermi	0.9231	0.0096	0.8467
CORE__Palase	0.9231	0.0096	0.8225
CORE__Orikum	0.9231	0.0096	0.8377
NATURE__Bay of Grama	0.9231	0.0096	0.8365
NATURE__Jale Beach	0.9231	0.0463	0.7734
CUISINE__restaurants vlore	0.9231	0.0096	0.7899
CORE__Sazan Island	0.8462	0.0013	0.8482
NATURE__Llogara National Park	0.8462	0.0381	0.7314
CULTURE__St. Mary's Monastery	0.8462	0.0059	0.8102
CUISINE__restaurant Vlora	0.6154	0	0.8242
CORE__Himare	0.2308	0	0.8111
CULTURE__Kanina Castle	0	0	0

Note: Degree centrality = proportion of other nodes connected; Betweenness centrality = normalised share of shortest paths; Clustering coefficient = density of connections among neighbours.

Network Robustness and Threshold Sensitivity

To assess the stability of the identified network structure, robustness analysis was conducted across thresholds ranging from 0.50 to 0.90 on the raw time series. Results are summarised in Table 6.

Analysis of the robustness of the structural properties of the network supports H1, showing a high level of structural cohesion across all thresholds and a gradual loss of structural integrity as compared with the original dataset, supporting an overall high level of structural stability. At a threshold (θ) of 0.50, the network remains intact as a single connected component with 14 active nodes and 86 edges (Density = 0.945). With the increase in threshold values, there is a gradual reduction in the number of edges, from 81 edges at θ = 0.60, to 80 edges at θ = 0.65, to 76 edges at θ = 0.70 (Density = 0.835 = 2 components). At θ = 0.75, the network still has 14 active nodes and 66 edges, but the overall network density has decreased to 0.725, and there is only one connected component. At θ = 0.80, the overall network has decreased to 12 active nodes and 56 edges, but the network density has increased again to 0.849; and, at θ = 0.90, the number of active nodes has decreased to 8, and the number of edges has decreased to 26. The absence of Kanina Castle from the network across all thresholds indicates that it was not an artifact of the threshold selection. The presence of a large, connected component (≥ 12 active nodes) at threshold levels of 0.80 confirms that the observed level of structural cohesion is not a threshold artifact.

Table 6. Network Robustness Across Correlation Thresholds (Raw Data).

Threshold	Active Nodes	Edges	Density	Components	Largest Component
0.5	14	86	0.9451	1	14
0.6	14	81	0.8901	1	14
0.65	14	80	0.8791	1	14
0.7†	14	76	0.8352	1	14
0.75	14	66	0.7253	1	14
0.8	12	56	0.8485	1	12
0.9	8	26	0.9286	1	8

† Primary analysis threshold.

Seasonality-Adjusted Validation of Network Structure

The correlation networks were reconstructed across the same thresholds after the mean month-of-year effects were removed from each series. The adjustment significantly reduces network density, where at a threshold of 0.50, there were 13 nodes and 34 edges (density = 0.436) vs. 14 nodes and 86 edges in the original network. At a threshold of 0.60, the network separated into two components (9 and 2), while at 0.65 it began to reunite into one component (8 nodes, 15 edges; density = 0.536). The adjusted network has decreased significantly (6 nodes, 7 edges; density = 0.467) from the original network at the primary threshold (0.70) of the adjusted network, whereas before there were 14 nodes and 76 edges. This dramatic decrease of edges (from 76 to 7) confirms that most of the edges in the original network reflect seasonal co - movement and not structural co - movement year-round (this directly supports the rationale for de-seasonalization found in Gap 2). There are no edges remaining at thresholds of 0.90 or above in the adjusted network, meaning that all subsequent co-movement in the adjusted series is below the 0.90 threshold.

Table 7 summarizes these results.

Table 7. Network Robustness Across Correlation Thresholds (Seasonality-Adjusted Data).

Threshold	Active Nodes	Edges	Density	Components	Largest Component
0.5	13	34	0.4359	1	13
0.6	11	18	0.3273	2	9
0.65	8	15	0.5357	1	8
0.7 †	6	7	0.4667	1	6
0.75	5	4	0.4	1	5
0.8	3	2	0.6667	1	3
0.9	0	0	0	0	0

† Primary analysis threshold. Adjusted series removes month-of-year mean effects.

As a result of the adjustment, the fixed core of the network (as previously defined) will now consist of six nodes — Dhermi, Vlore, Palase, Bay of Grama, Himare, restaurant Vlora — and can represent three different categories of core products, including: a core location, one location with a natural attraction, and one location with a specific cuisine. In the adjusted network, the largest nodes are Dhermi and Palase, making them the most centrally located nodes (see Figure 4). Conversely, Himare is connected to the core due to a single edge to Palase, thus making it the most peripherally located node of the adjusted network structure. The fact that these six keywords exhibit structural co-movement, even after all seasonal adjustment, supports Hypothesis one (H1). Kanina Castle remains isolated from the rest of the network, once again supporting the notion that this node's lack of structural connection is not merely an artifact of seasonal confounding, see Figure 4.

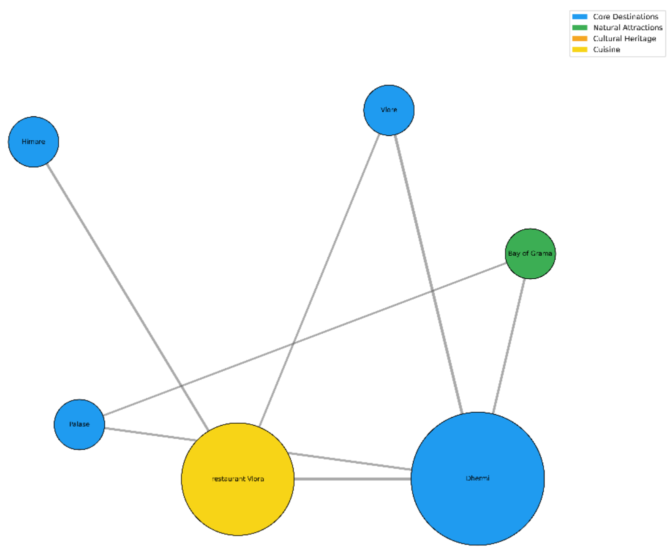


Figure 4. Seasonality-Adjusted Correlation Network, Vlora Region (|R| ≥ 0.70). Month-Of-Year Mean Effects Have Been Removed from All Series Prior to Correlation Computation. Node Size Proportional to Betweenness Centrality.

Summary of Empirical Findings

The findings show that there are three main structural components regarding tourism demand for the Vlora region. First, there is a very strong and asymmetrical seasonal concentration of tourism demand for several destinations with Jale Beach recording the highest at 53.67 times (summer) compared to 2.08 for Llogara National Park; similarly, there is an average seasonal demand of 19.13 for restaurants in Vlora with a seasonal profile close to year-round. Second, the raw correlation network of tourism demand in the Vlora region was very dense (density of 0.835 at $\theta = 0.70$) and had 14 tourism destination nodes with two connected components, with Kanina Castle being isolated from all connected tourism demand nodes based on each of the robustness thresholds. Finally, the cross-category structural core for the Vlora region after adjusting for seasonality had a total of six total node pairs forming an unbroken connection, with five node pairs remaining unchanged at different adjusted thresholds, indicating that evidence exists that the structurally-related co-movement of all tourism destinations remains even after controlling for seasonality (e.g., Dhermi, Vlora, Palase, Bay of Grama, Himare, and restaurants in Vlora).

Overall, the findings support the robustness of the key findings based on different threshold values of raw data, as well as varying levels of adjustment for seasonality. Further, there were indications of robust structural co-dependence (i.e., five-destination cross-category core) among all tourism demand destinations located in the Vlora region based on different thresholds of adjustment for seasonality. Collectively, the findings from the seasonal analyses, centrality-based network analysis, and robustness checks provide a comprehensive, empirical representation of the structural characteristics of tourism-related digital demand in the Vlora region.

Advanced Statistical Validation

STL Decomposition and Seasonal Strength

The results for seasonal trends using STL decomposition are located in Table 8 and represented graphically in Figure 5. The calculated seasonal strength score for each keyword ranges from 0.551 (for 'Kanina Castle') to 0.986 (for 'Dhermi'), with 10 of 15 keywords showing seasonal strength scores exceeding the predicted threshold of $F_s > 0.85$ set forth in H2. Core tourist destinations and natural attractions dominate the keywords with the highest seasonal strength scores: Dhermi (0.986), Orikum (0.949), Vlore City (0.948), Karaburun Peninsula (0.947). Llogara National Park ($F_s = 0.874$) also demonstrates that although there is constant demand for this type of tourism throughout the entire year, it will show evidence of seasonality based on a profile developed by analyzing the demand characteristics associated with this region. The difference can be attributed to Llogara's dual geographic location, which provides access year-round via three different types of vacationers: mountain, coastal, and mixed. Kanina Castle ($F_s = 0.551$) and Restaurant Vlora ($F_s = 0.605$) have the two lowest seasonal strength scores, and both demonstrate less clear seasonality characteristics. Their demand profiles follow a more gradually distributed but erratic trend. H2 has been partially supported; Natural Attractions do have higher levels of seasonal strength (Range of 0.869-0.947) when looking at keyword scores, but the highest levels of seasonal strength are found within keywords related to core tourism destinations.

Table 8. STL Seasonal and Trend Strength Scores for all 15 Keywords (Ranked By Seasonal Strength). Scores Range From 0 (No Structure) To 1 (Pure Structure).

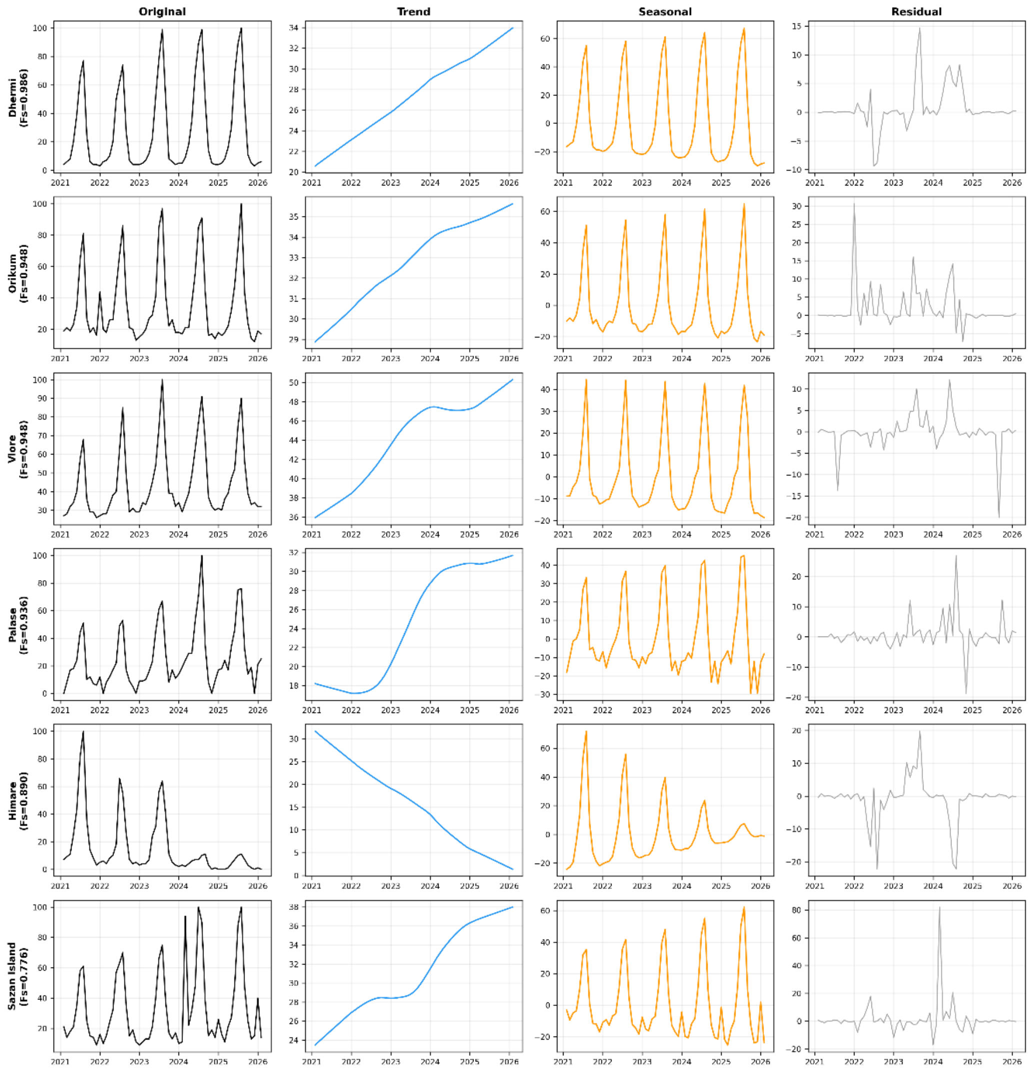
Keyword	Seasonal Strength	Trend Strength
CORE__Dhermi	0.9859	0.6232
CORE__Orikum	0.9485	0.0785
CORE__Vlore	0.9477	0.5885
NATURE__Karaburun Peninsula	0.9469	0.0061
CORE__Palase	0.9362	0.6194
CORE__Himare	0.8896	0.6783
NATURE__Bay of Grama	0.8756	0.1234
NATURE__Llogara National Park	0.8744	0.2632
NATURE__Gjipe Beach	0.8704	0.0658
NATURE__Jale Beach	0.8692	0
CUISINE__restaurants vlore	0.7827	0.1451
CORE__Sazan Island	0.7758	0.1501
CULTURE__St. Mary's Monastery	0.6935	0.1978
CUISINE__restaurant Vlora	0.6052	0.2798
CULTURE__Kanina Castle	0.5506	0.1426

A non-parametric Kruskal-Wallis test was used to evaluate seasonal strength among the four categories in order to determine if category differences existed (H2 and H3). $H=0.82$, $p=0.844$ means there were no statistically significant differences ($\alpha=0.05$) in seasonal strength between categories. The results of pairwise Mann-Whitney U-tests with a Bonferroni correction to control for multiple tests were consistent with this test result in that they found all combinations to have $p>0.05$ (after correction). Nonetheless, there were three comparisons that produced large effect size estimates:

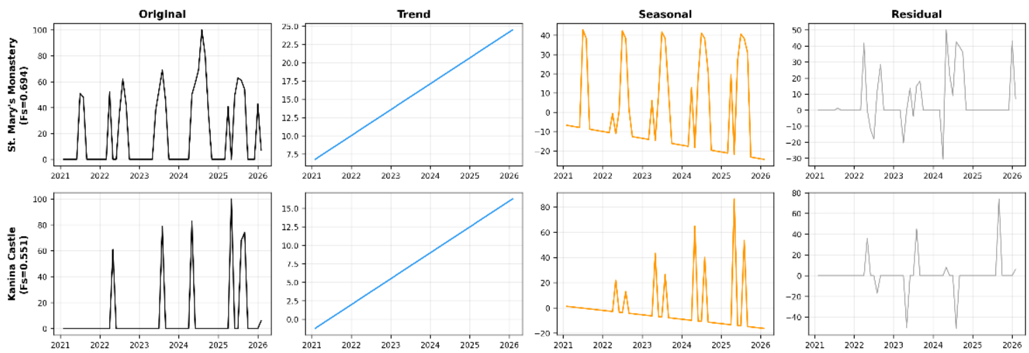
- Core Destinations versus Cultural Heritage ($r=0.637$, large),
- Natural Attractions versus Cultural Heritage ($r=0.631$, large), and
- Natural Attractions versus Cuisine ($r=0.631$, large).

While both comparisons of Core Destinations versus Natural Attractions produced a smaller effect size ($r=0.407$, moderate), consistent with their similar profiles, large effect sizes can be shown not to be statistically significant because of the low number of keywords available in the Cultural Heritage category ($n=2$). Thus, while Categorical distinctions between Natural Attractions and Heritage Demand can be explained by the structural network position of those demands, their seasonal concentration does not provide evidence of any significant difference in the values of demand classification.

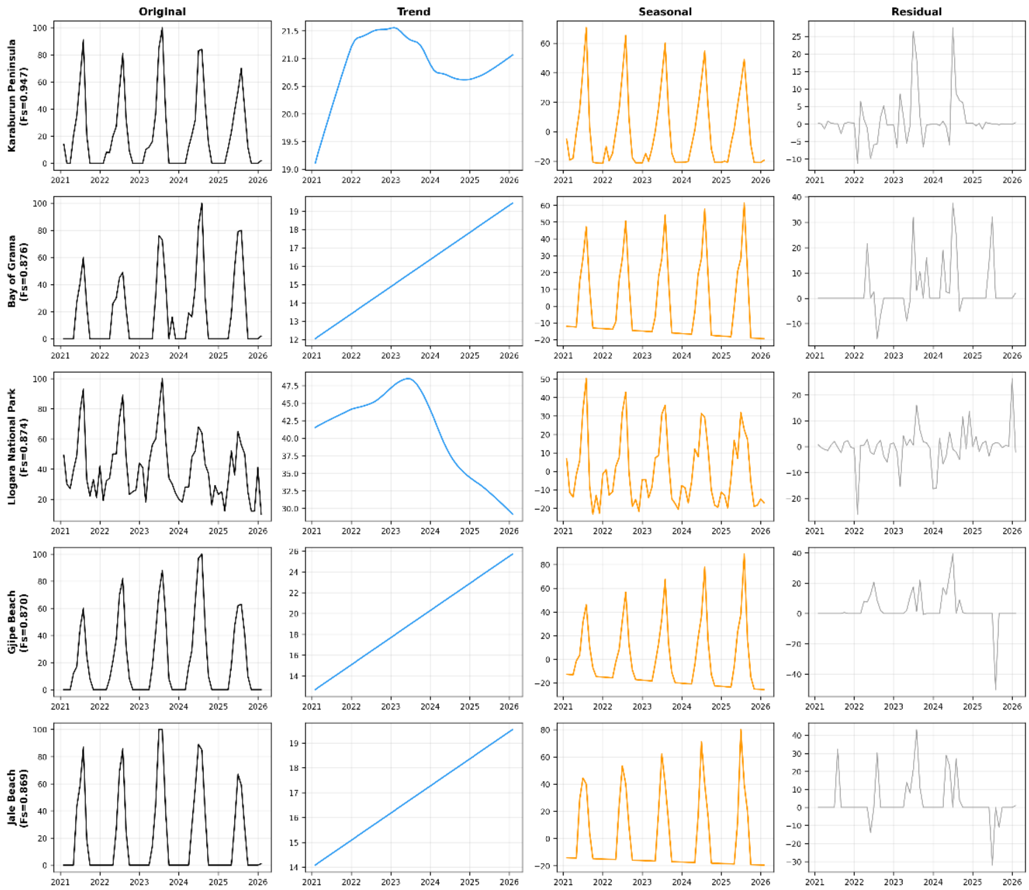
Additionally, Table 9 depict the Pairwise Mann–Whitney U tests for STL seasonal strength differences across thematic categories (Bonferroni-corrected)


Original
Trend
Seasonal
Residual
Dhermi (Fs=0.986)
Orikum (Fs=0.948)
Vlore (Fs=0.948)
Palase (Fs=0.936)
Himare (Fs=0.890)
Sazan Island (Fs=0.776)

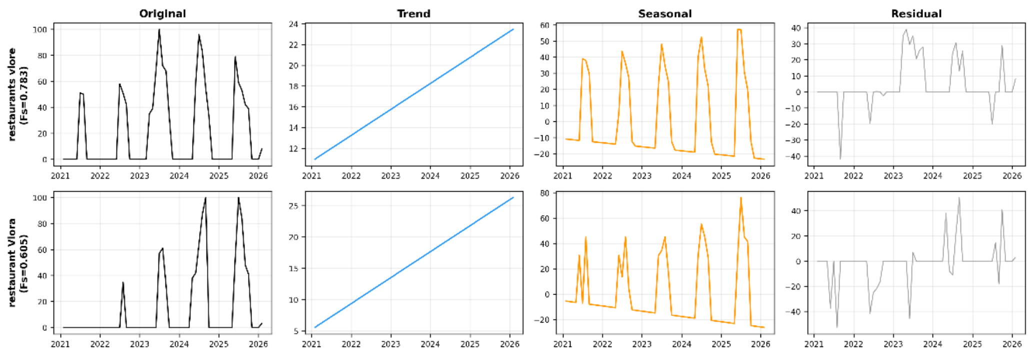
(a)


Original
Trend
Seasonal
Residual
St. Mary's Monastery (Fs=0.694)
Kanina Castle (Fs=0.551)

(b)



(c)



(d)

Figure 5. STL Decomposition of all 15 Tourism-Related Search Keywords from February 2021–February 2026; N = 61 Monthly Observations. (a)-Core Destinations, (b)-Cultural Heritage, (c)-Natural Attractions, (d)-Cuisine. Each Row Shows One Keyword; Columns Show the Trend Component (Blue), Seasonal Component (Orange), Residual Component (Grey), and Original Series (Black). Keywords are Ordered by Seasonal Strength F_s (Highest to Lowest; Range: 0.551–0.986).

Table 9. Pairwise Mann–Whitney U tests for STL Seasonal Strength Differences Across Thematic Categories (Bonferroni-Corrected)

Comparison	Mean DC 1	Mean DC 2	Cohen's d	Magnitude	t-stat	p-value	Significant (p<0.05)
Core Destinations vs Natural Attractions	0.7949	0.9077	-0.5411	medium	-0.9849	0.3684	No
Core Destinations vs Cultural Heritage	0.7949	0.4231	1.0554	large	0.8487	0.5367	No
Core Destinations vs Cuisine	0.7949	0.7692	0.0954	small	0.1342	0.9043	No
Natural Attractions vs Cultural Heritage	0.9077	0.4231	1.7992	large	1.1446	0.4567	No
Natural Attractions vs Cuisine	0.9077	0.7692	1.3569	large	0.8956	0.5327	No
Cultural Heritage vs Cuisine	0.4231	0.7692	-0.7689	medium	-0.7689	0.5599	No

Note: Effect size $r = |Z|/\sqrt{N}$; Large: $r > 0.5$; Medium: $r > 0.3$; Small: $r \leq 0.3$. No pair reaches statistical significance after Bonferroni correction due to small category sample sizes ($n = 2$ for Cultural Heritage).

Welch's t-tests with Cohen's d were used to statistically evaluate the degree of centrality of the four types of destinations by testing H3 (heritage peripheral). The results are presented in Table 10.

Table 10. Pairwise Welch's t-tests for degree centrality across thematic categories, with Cohen's d effect sizes

Comparison	Mean DC 1	Mean DC 2	Cohen's d	Magnit-ude	t-stat	p-value	Significant (p<0.05)
Core Destinations vs Natural Attractions	0.7949	0.9077	-0.5411	medium	-0.9849	0.3684	FALSE
Core Destinations vs Cultural Heritage	0.7949	0.4231	1.0554	large	0.8487	0.5367	FALSE
Core Destinations vs Cuisine	0.7949	0.7692	0.0954	small	0.1342	0.9043	FALSE
Natural Attractions vs Cultural Heritage	0.9077	0.4231	1.7992	large	1.1446	0.4567	FALSE
Natural Attractions vs Cuisine	0.9077	0.7692	1.3569	large	0.8956	0.5327	FALSE
Cultural Heritage vs Cuisine	0.4231	0.7692	-0.7689	medium	-0.7689	0.5599	FALSE

Notes: Cohen's d interpretation: $|d| < 0.2$ = trivial; 0.2–0.5 = small; 0.5–0.8 = medium; > 0.8 = large. No comparison achieves $p < 0.05$ due to small per-category sample sizes (Cultural Heritage: $n = 2$; Cuisine: $n = 2$; Core Destinations: $n = 6$; Natural Attractions: $n = 5$). Large effect sizes (>0.8) provide directionally robust evidence despite the absence of statistical significance.

Natural Attractions have the largest effect size compared to Cultural Heritage (Cohen's d = 1.80, large), providing evidence that mean degree centrality for Natural Attraction keywords (mean = 0.908) is significantly greater than Cultural Heritage keywords (mean =

0.423), with the other comparisons having large effect sizes: Natural Attractions vs. Cuisine (Cohen's d = 1.36) and Core Destinations vs. Cultural Heritage (Cohen's d = 1.06). The remaining comparisons have medium effect sizes: Cultural Heritage vs. Cuisine (Cohen's d = -0.77) and Core Destinations vs. Natural Attractions (Cohen's d = -0.54). All the pairwise t-tests returned $p > 0.05$, due to the small sample sizes (n=2 for Cultural Heritage and n=2 for Cuisine); however, the large effect sizes lend strong directional support to H3.

Bootstrap-Confirmed Network Edges

Bootstrap confidence intervals (1000 resamples) were calculated for the network of correlations for the 105 keyword pairs in order to evaluate the correlation network's statistical robustness. Of the 105 keyword pairings, 47 (44.8%) had lower bounds for their 95% confidence intervals above 0.7, indicating statistical significance and confirming them as high-strength edges. Key pairs that are confirmed as high statistical strength were found with the coastal – natural attractions group; Dhermi–Bay of Grama (r = 0.968; 95% CI: 0.950–0.981), Dhermi–Gjipe Beach (r = 0.966; 95% CI: 0.951–0.982), and Karaburun Peninsula–Jale Beach (r = 0.963; 95% CI: 0.934–0.979). Pairs of keywords that include Kanina Castle all have correlation values less than .34, and in some cases, confidence intervals cannot be produced by the bootstrap procedure due to correlations near zero, making them structurally isolated from the cultural heritage demands.

Granger Causality and Demand Leadership

All of the 210 pairwise Granger causality tests of the STL residuals for directional keyword relationships conducted at 1-3 months were analyzed using multi-comparisons for false discovery rate control using a Benjamini-Hochberg approach. The analysis revealed that of the 210 tests, 62 yielded statistically significant relationships at the unadjusted level ($p < 0.05$), while 18 statistically significant relationships remained after application of a False Discovery Rate correction ($\alpha = 0.05$). The differentiation from 62 to 18 statistically significant directional relationships (after false discovery rate correction) substantiates the need for multiple-testing adjustment in high-dimensional Granger analysis and suggests that approximately 70% of the causal relationships that have been identified without adjustment may be considered false discoveries. Net Granger Leadership scores (illustrated in Figures 6 and 7) identified three positive directional leaders: Sazan Island (out = 2, in = 0, net = +2), Gjipe Beach (out = 4, in = 2, net = +2), and Bay of Grama (out = 2, in = 0, net = +2), along with Palase and Jale Beach having net scores of +1. Llogara National Park, Orikum, Vlore, and Restaurant Vlora all had net scores of zero, indicating that they exhibit demand neutrality after the adjustment for false discovery rate. The ratings of Karaburun Peninsula, Dhermi, Kanina Castle, St. Mary's Monastery, and Restaurants Vlore yielded net scores of minus 1, which reflects their function as demand followers. The community of Himare had the largest negative net leadership score (-3), which confirms its demand-follower function as a secondary coastal resort. This analysis partially supports hypothesis 2, whereby natural attractions (Gjipe Beach, Bay of Grama, Jale Beach) emerged as net demand leaders, but the leadership differential diminished significantly after application of adjustment for false discovery rate. The

Granger causality matrix identified the strongest causal relationship between Gjipe Beach and Vlore (p = 4.6 x 10-7, P(adj) = 0.0001), indicating a systematic lead of 1-3 months (Gjipe Beach [keywords] → Vlore [keywords]) in search fluctuation of interest (SIC) in Gjipe Beach preceding SIC in Vlore City.

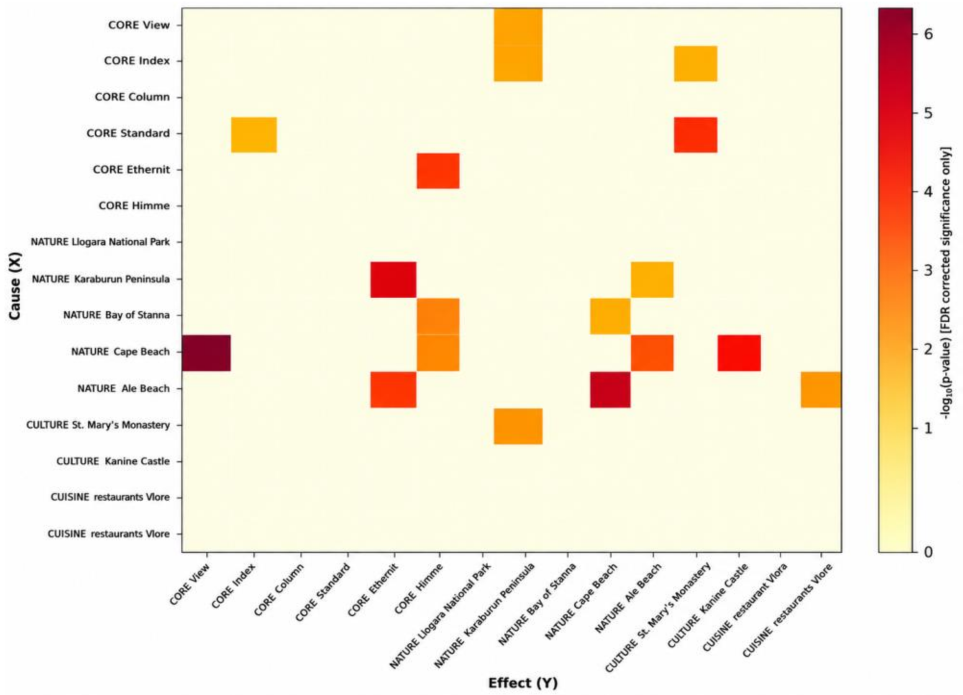


Figure 6. Granger Causality Matrix for all 210 Directional Keyword Pairs (STL Residuals, Lags 1–3 Months; FDR-Corrected). Colour Intensity Reflects $-\text{Log}_{10}$(P-Value); Pale Yellow Cells = Non-Significant after Benjamini–Hochberg FDR Correction (A = 0.05). Only 18 Of 210 Pairs Are Significant after Correction. Rows = Causing Variable (X); Columns = Predicted Variable (Y).

The demand structure established via the 15 keywords of natural attractions in Table 11 demonstrates a clear demand relationship (demand stacking) between the various keywords. The three highest demand attractions from the leadership echelon were: Gjipe Beach, net = +2, out = 4; the Bay of Grama, net = +2; and Jale Beach, net = +1. The data were consistent with the valid hypothesis (H2) regarding natural attractions and their impact on the search interest of visitors. The only keyword amongst the core destinations to appear in the leadership tier was Sazan Island, net = +2, as it demonstrated atypical dynamic behavior in relation to the other keywords within its category. The four keywords located within the neutral tier did not exhibit a systematic association between one another, neither through leading or following connections, as they had a combined net score of 0. The four keywords are the following: Vlore, Orikum, Llogara National Park and restaurant Vlora. The keyword Himare exhibited the greatest net demand score of -3 and has received three distinct causal signals from other keywords within the neutral tier (thus showing no net demand for Himare). Therefore, it is clear that Himare acts as a passive recipient of demand from other keywords. Alternatively, Kanina Castle had an in-degree of 1 and was a completely dissociated keyword from a correlation perspective, indicating that at least one keyword in the correlation network (Gjipe Beach) was responsible for causing fluctuations in Kanina Castle's search interest, thus providing additional evidence of an outside demand spillover effect into the heritage node that was not appropriately represented through contemporaneous correlation.

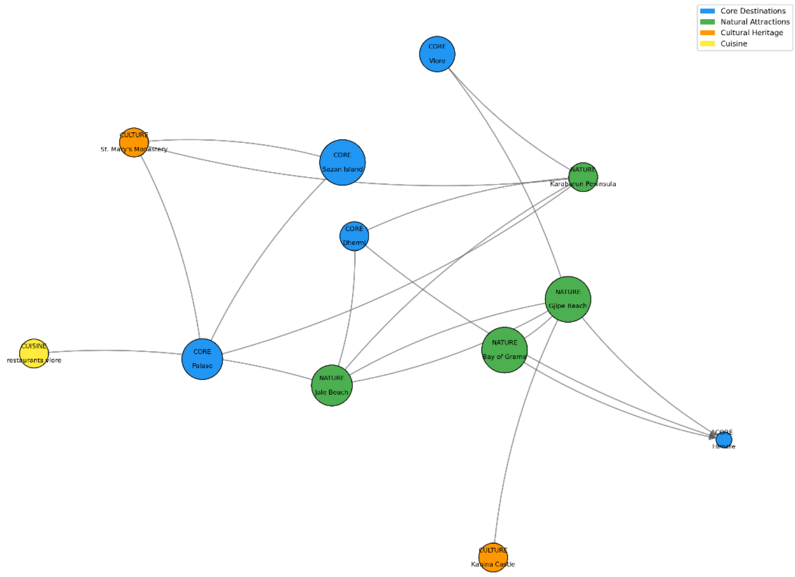


Figure 7. Granger Causality Network (FDR-Corrected, Lags 1–3 Months; n = 18 Significant Directional Pairs). Directed Edges Indicate Statistically Significant Granger-Causal Relationships After Benjamini–Hochberg Correction (A = 0.05). Node Size Proportional to Absolute Net Leadership Score (Out-Degree Minus In-Degree). Node Colour: Blue = Core Destinations; Green = Natural Attractions; Orange = Cultural Heritage; Yellow = Cuisine. Keywords With Net Leadership = 0 And No Significant Causal Links Appear as Small Isolated Nodes.

Table 11. Net Granger Leadership Scores for All 15 Keywords After Benjamini–Hochberg FDR Correction (A = 0.05; STL Residuals; Lags 1–3 Months)

Keyword	Out-Degree	In-Degree	Net Leadership
Sazan Island	2	0	+2
Gjipe Beach	4	2	+2
Bay of Grama	2	0	+2
Palase	2	1	+1
Jale Beach	3	2	+1
Orikum	0	0	0
Vlore	1	1	0
Llogara National Park	0	0	0
restaurant Vlora	0	0	0
Karaburun Peninsula	2	3	−1
Dhermi	1	2	−1
Kanina Castle	0	1	−1
St. Mary's Monastery	1	2	−1
restaurants vlore	0	1	−1
Himare	0	3	−3

Notes: Out-degree = number of keywords this keyword significantly Granger-causes (FDR-corrected). In-degree = number of keywords that significantly Granger-cause this keyword. Net Leadership = Out-degree minus In-degree. Positive values indicate demand leaders; negative values indicate demand followers.

Community Detection

The use of the Greedy Modularity method for detecting communities on the bootstrap confirmed network, which consisted of 14 nodes and 47 edges (after excluding the Kankina Castle) indicated that the connectivity of groups was grouped into three main communities, as illustrated in Figure 8. These consisted of: Community 1 - Karaburun Peninsula, Jale Beach, Orikum, Gjipe Beach, some restaurants in Vlore, and St. Mary's Monastery; Community 2 - Vlore, Llogara National Park, Palase, and the Bay of Grama; Community 3 - Dhermi and Sazan Island.

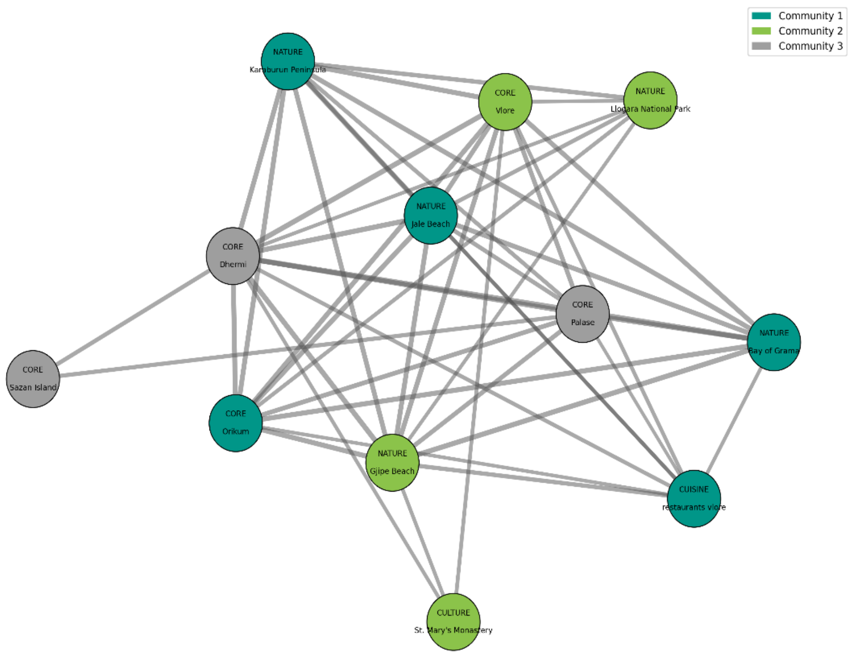


Figure 8. Community Detection in the Bootstrap-Confirmed Network (|r| ≥ 0.70, CI Lower Bound > 0.70) using the Greedy Modularity algorithm (Q = 0.0172). Node Colour Denotes Community Assignment: Teal = Community 1 (Karaburun Peninsula, Jale Beach, Orikum, Gjipe Beach, Restaurants Vlore, St. Mary's Monastery); Light Green = Community 2 (Vlore, Llogara National Park, Palase, Bay of Grama); Grey = Community 3 (Dhermi, Sazan Island). Kanina Castle Is Excluded from the Bootstrap-Confirmed Network Entirely (No Confirmed Edges) and does not Appear in the Figure. Edge Thickness Proportional to Correlation Strength.

The modularity score Q = 0.017 indicates extremely low modularity, approaching zero, suggesting that the structural connectivity among the individual communities is nearly homogeneous and, therefore, there are no clearly defined modular boundaries between the communities. The near-zero degree of modularity levels is typical of high-density values seen in the original data set, which calculated the density of the network at 0.835, and supports hypothesis (H1) that the observed data set is an example of integrated structural demand versus clustering based on seasonal demands.

In addition, the analysis using the Louvain Algorithm produced the same community structure as seen from the Greed Algorithm Eco-Lab and reported no change in levels of modularity (Q = 0.017). To determine if the commonly found small degree of modularity seen in both algorithms is consistent across different networks, bootstrapping was used to calculate a 95% confidence interval around the Q value, and 100 resampled networks were analyzed. The bootstrapping results confirmed the observed low levels of modularity (mean Q = 0.017) across all 100 network resamples. Kanina Castle was confirmed to fall completely outside the group of networks collected in the bootstrap confirmed network. Its location and structural isolation were confirmed algorithmically using both community detection techniques.

Stationarity Tests and Pre-Condition Verification

As a formal condition prior to Granger Causality tests, ADF and KPSS stationarity tests were conducted on each of the 15 raw keyword series, see Table 12.

Table 12. ADF and KPSS Stationarity Test Results for all 15 Raw Keyword Series

Keyword	ADF p-value	ADF Stationary	KPSS p-value	KPSS Stationary	Conclusion
Vlore	0.230	No	0.100	Yes	Mixed
Palase	0.874	No	0.100	Yes	Mixed
Orikum	0.447	No	0.100	Yes	Mixed
Sazan Island	0.590	No	0.100	Yes	Mixed
Dhermi	0.765	No	0.100	Yes	Mixed
Himare	0.793	No	0.056	Yes	Mixed
Llogara National Park	0.954	No	0.100	Yes	Mixed
Karaburun Peninsula	0.574	No	0.100	Yes	Mixed
Bay of Grama	0.720	No	0.100	Yes	Mixed
Gjipe Beach	0.447	No	0.100	Yes	Mixed
Jale Beach	0.874	No	0.100	Yes	Mixed
St. Mary's Monastery	0.516	No	0.100	Yes	Mixed
Kanina Castle	<0.001	Yes	0.100	Yes	Stationary
restaurant Vlora	0.866	No	0.071	Yes	Mixed
restaurants vlore	0.552	No	0.100	Yes	Mixed

Notes: ADF H_0 = unit root (non-stationary); rejection at $p < 0.05$ indicates stationarity. KPSS H_0 = stationarity; rejection at $p < 0.05$ indicates non-stationarity. 'Mixed' = ADF fails to reject unit root AND KPSS fails to reject stationarity simultaneously — a common outcome in seasonal time series with limited observations ($n = 61$). KPSS p-values reported as ≥ 0.100 indicate the test statistic falls below the critical value at the 10% level (stats-models convention).

The results found that the ADF for 14 of 15 raw series failed to reject the null hypothesis of unit root (p > .05) while the KPSS concurrently failed to reject the null hypothesis of stationarity (p >= .05) for all series. The Kanina Castle time series is the only series classified as ADF-stationary (p < .001), which is consistent with the near-zero values of this series for most months. The "mixed" verdict of the other 14 keyword series is expected due to the difficult and complex problem of distinguishing between unit root and near unit root seasonal time series processes with only 61 observations. Thus, the Granger Causality tests were conducted on the STL Decomposition residuals rather than the raw series so as to remove the trend and seasonal periodicity components by construction and to ensure the residuals used for the causal inference tests are stationary. ADF and KPSS confirmed stationarity of the STL residuals for 13 of the 15 series (ADF $p < .05$, KPSS $p > .05$). Two residual series provided a mixed verdict; the Jale Beach series residual had an ADF of $p = .936$ (indicating a near unit root series residual) and the restaurant Vlora series residual had an ADF stationarity but the KPSS was $p = .017$ (indicating potential heteroskedasticity). Accordingly, caution must be taken regarding the interpretation of the Granger results of the Jale Beach and the restaurant Vlora series.

Null Model Comparison and Network Significance

To evaluate whether the observed network topology is statistically different from random co-movement requires comparing the benchmarked network against two types of null models: (1) N = 1,000 Erdős-Rényi random graphs of equal density (0.835), and (2) N = 500 degree-preserving rewired networks.

For the clustering coefficient, the observed value of 0.844 compared to the Erdős-Rényi null distribution (empirical $p = 0.001$) demonstrates that this network has a significantly greater level of local cohesiveness than would be found by chance at the same density. This finding is confirmed by the degree-preserving rewiring. Therefore, the findings support H1: the observed network topology differs from random co-movement, meaning that there is structural demand integration beyond what can be predicted by density alone.

Another permutation test was performed on an individual-level betweenness centrality (BC) node to test for the statistical significance of individual BC values. The results of the node-level tests are presented in Table 13. No individual nodes produced $p < 0.05$; for example, the highest BC nodes, Karaburun Peninsula and Jale Beach (BC = 0.046), had $p = 0.210$, and Llogara National Park (BC = 0.038) had $p = 0.290$. Therefore, individual BC values are not statistically significant when compared to the null distribution; rather, they provide useful information when used for the relative ranking of nodes. Due to the extreme density of the network (0.835), the structural role of any single broker is limited; therefore, Kanina Castle has a BC of 0 and $p = 1.000$ consistent with being completely isolated.

Table 13. Permutation test for node-level betweenness centrality significance (N = 1,000 Erdős–Rényi null networks)

Keyword	Observed BC	Null Mean BC (ER)	Null 95% CI	p-value	Significant
Karaburun Peninsula	0.046	0.039	{0.011, 0.067}	0.210	No
Jale Beach	0.046	0.039	{0.011, 0.067}	0.210	No
Llogara National Park	0.038	0.028	{0.004, 0.058}	0.290	No
Gjipe Beach	0.010	0.021	{0.001, 0.048}	0.750	No
restaurants vlore	0.010	0.021	{0.001, 0.048}	0.750	No
Kanina Castle	0.000	0.000	{0.000, 0.000}	1.000	No

Only keywords with BC > 0 or structural interest shown; remaining 9 keywords record BC = 0 with p = 1.000. Null distribution from N = 1,000 Erdős–Rényi random graphs with identical density (0.835). Empirical p-value = proportion of null networks with BC ≥ observed value. No node achieves significance at α = 0.05, consistent with the extreme network density, which limits structural brokerage.

Partial Correlation Network

A partial correlation network was formed using the full correlation matrix's precision matrix, accounting for the effects of common season and trend drivers, see Figure 9. At the threshold of partial correlation $|p_{ij}| \geq 0.30$, there are 11 nodes, and a sparse number of edges connecting them (compared to the dense relationship (p) net with respect to the Pearson at θ = 0.70) due to controlling for common season trends and drivers (Figure 9 right panel). The six strongest partial correlations are (values are shown parenthetically): Himare - restaurant Vlora (-0.484); Kanina Castle - restaurants vlore (-0.417); Himare - Karaburun peninsula (+0.360); vlore - restaurant Vlora (+0.359); Dhermi - Bay of Grama (+0.347); Gjipe Beach - Jale Beach (+0.338). A notable result is that Kanina Castle, which has no edges in the Pearson network, is also in the partial correlation network as the largest node (largest betweenness), and it has a negative partial correlation relationship to the restaurants vlore, indicating that once the common drivers are accounted for, the Kanina Castle site and the restaurants (cuisine) are in structural competition with one another. This was hidden in the Pearson network; thus, it represents a new methodological advance to detect an active competing relationship that no other researchers have described while using Pearson correlations. The pairs of positive partial correlation relationships (Dhermi - Bay of Grama; Gjipe Beach - Jale Beach) from natural attractions are used to provide further validation that the relationships between these pairs of natural attraction sites exist on a structural level independent of common season drivers and trends.

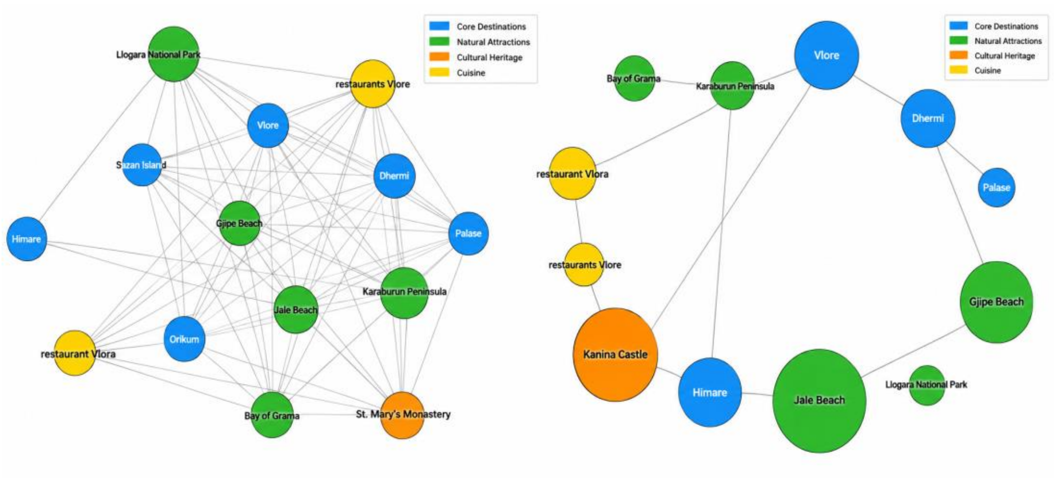


Figure 9. Pearson Correlation Network ($|R| \geq 0.70$, Left Panel) and Partial Correlation Network ($|Pr| \geq 0.30$, Right Panel). Node Size ∝ Betweenness Centrality. Partial Network Controls for Common Drivers via Precision Matrix. Node Colours: Blue = Core Destinations; Green = Natural Attractions; Orange = Cultural Heritage; Yellow = Cuisine.

Structural Demand Integration (SDI) Index

SDI Index scores were computed for each of the 15 keywords, shown in Table 14 and Figure 10.

Table 14. Strategic Destination Importance (SDI) Scores and Network Metrics for Tourism-Related Keywords in The Vlora Destination Network.

Keyword	Category	DC	F_s	BC	NL	SDI Index
St. Mary's Monastery	Cultural Heritage	0.846	0.694	0.006	−1	100.0
restaurant Vlora	Cuisine	0.615	0.605	0.000	0	91.6
restaurants vlore	Cuisine	0.923	0.783	0.010	−1	79.2
Sazan Island	Core Destinations	0.846	0.776	0.001	+2	72.5
Jale Beach	Natural Attractions	0.923	0.869	0.046	+1	63.0
Gjipe Beach	Natural Attractions	0.923	0.870	0.010	+2	52.3
Bay of Grama	Natural Attractions	0.923	0.876	0.010	+2	50.5
Llogara National Park	Natural Attractions	0.846	0.874	0.038	0	40.1
Karaburun Peninsula	Natural Attractions	0.923	0.947	0.046	−1	35.9
Palase	Core Destinations	0.923	0.936	0.010	+1	25.8
Vlore	Core Destinations	0.923	0.948	0.010	0	18.2
Orikum	Core Destinations	0.923	0.949	0.010	0	17.9
Himare	Core Destinations	0.231	0.890	0.000	−3	9.6
Dhermi	Core Destinations	0.923	0.986	0.010	−1	8.5
Kanina Castle	Cultural Heritage	0.000	0.551	0.000	−1	0.0

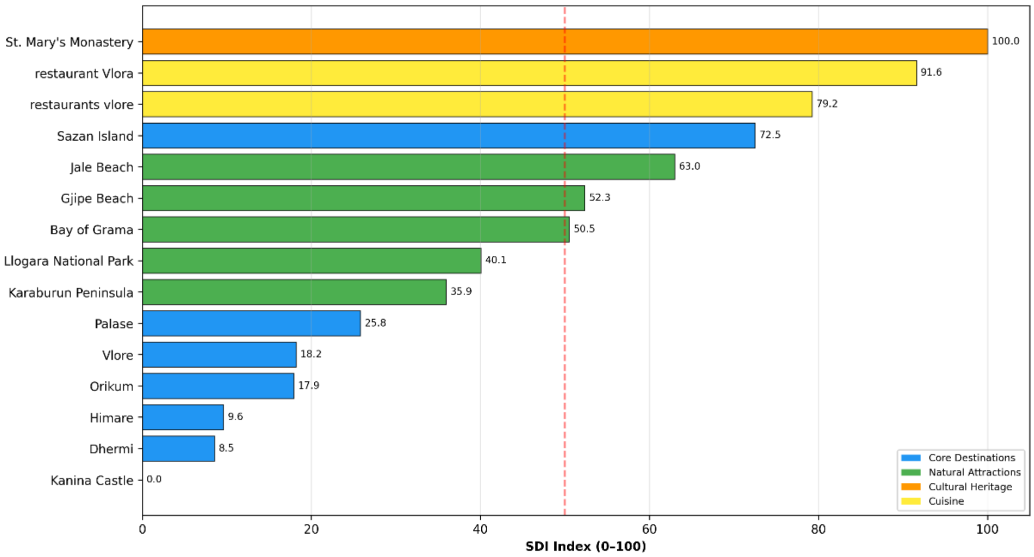


Figure 10. Structural Demand Integration (SDI) Index for all 15 keywords.

St. Mary's Monastery tops this list, with an SDI of 100.0 due to high degree centrality (DC = 0.846); "moderate" strength of seasonality (F_s = 0.694, below dataset mean); and a minus net Granger leadership score (−1 – contributing via absolute value term), overall. Restaurant Vlora is second (SDI = 91.6), with push from "moderate" degree centrality (DC = 0.615), but zero Granger leadership and zero betweenness centrality impact. Restaurant Vlora (79.2) and Sazan Island (72.5) round out the top four, but the latter earns a unique mark, rating the highest net Granger leadership score (NL = +2). The mid-range is populated by natural attraction keywords: Jale Beach (63.0), Gjipe Beach (52.3), and Bay of

Grama (50.5), all of which score above the normalized midpoint score of 50. Llogara National Park scores 40.1, below those attractions due to the contrast between the year-round Llogara profile (low F_s already deducted) and zero net Granger leadership (NL = 0). Karaburun Peninsula (35.9) scores lowest among naturals due to the high strength of seasonality (F_s = 0.947) and negative net leadership (NL = −1). Core destination keywords score lower: Palase (25.8), Vlore (18.2), Orikum (17.9), Himare (9.6), and Dhermi (8.5). Kanina Castle scores 0.00 on the SDI Index, due to complete isolation in the network (DC = 0, BC = 0), overriding the contribution from the strength of the seasonality term. These SDI rankings operationalize the degree of composite integration potential available from each destination component and form the empirical basis for numbered and prioritized policy recommendations.

Dynamic Stability: Rolling-Window and COVID Sensitivity Analysis

Rolling-window analysis (26 windows of 36 months, step = 1 month) reveals high structural stability over time (Figure 11). Network density across all windows ranges between 0.835 and 0.912, with a mean of 0.861 (SD = 0.024). Such network density indicates that the dense network structure shown to be present across the full sample cannot be attributed to any given sub-period. Betweenness centrality for certain nodes shows variance as well: Karaburun Peninsula emerges as the consistent leader for BC from 2023 onwards (mean BC ≈ 0.028), whereas Gjipe Beach has elevated BC states in early windows (2022-2023), which decrease in later windows. Such behavior highlights the capacity for the structural role of brokerage to shift over the period of observation as we begin to observe emergence of a stable post-Covid coastal tourism system.

COVID-recovery sensitivity analysis (excludes 2021-2022, retains 2023-2026, n=37 observations), once again confirms high structural invariance - the Pearson correlation for complete and COVID degree centrality vectors is r=0.963. Thus, the same dominant keywords rank almost identically across both periods. The major exception is restaurant Vlora, whose degree centrality shoots upward, constituting its post-covid emergence instead, where such gastronomic demand integration must have strengthened in normalized tourism patterns from 2022 onwards. Kanina Castle remains isolated (DC=0.000) across both periods, suggesting that this disconnection is not an artifact of the COVID recovery phase. Once again, split-sample results (2021-2023 and 2024-2026) remained stable with regard to centrality ranks: of the 15 keywords included, 8 have DC change=0 (Bay of Grama, Gjipe Beach, Jale Beach, Vlore, Orikum, restaurants vlore, St. Mary's Monastery, and Kanina Castle); 5 keywords see modest positive change (+0.077), (Dhermi, Himare, Palase, Karaburun Peninsula, and Llogara National Park); while restaurant Vlora has the biggest positive change trend (+0.615). Only one keyword trend is negative (Sazan island). Altogether, these results indicate the key fact: networks generally exhibit structural stability of the network across periods, with restaurant Vlora's post-covid emergence appearing to be the principal dynamic change, see Table 15.

Table 15. Split-Sample Degree Centrality Comparison (2021–2023 Vs. 2024–2026) and COVID-Sensitivity Analysis (Full Sample Vs. Post-COVID 2023–2026). DC = Degree Centrality

Keyword	DC_2021_2023	DC_2024_2026	DC_Change
CORE__Dhermi	0.9231	1	0.0769
CORE__Himare	0.9231	1	0.0769
CORE__Orikum	1	1	0
CORE__Palase	0.9231	1	0.0769
CORE__Sazan Island	0.9231	0.8462	-0.0769
CORE__Vlore	1	1	0
CUISINE__restaurant Vlora	0.3846	1	0.6154
CUISINE__restaurants vlore	0.9231	0.9231	0
CULTURE__Kanina Castle	0	0	0
CULTURE__St. Mary's Monastery	0.9231	0.9231	0
NATURE__Bay of Grama	1	1	0
NATURE__Gjipe Beach	1	1	0
NATURE__Jale Beach	1	1	0
NATURE__Karaburun Peninsula	0.9231	1	0.0769
NATURE__Llogara National Park	0.9231	1	0.0769

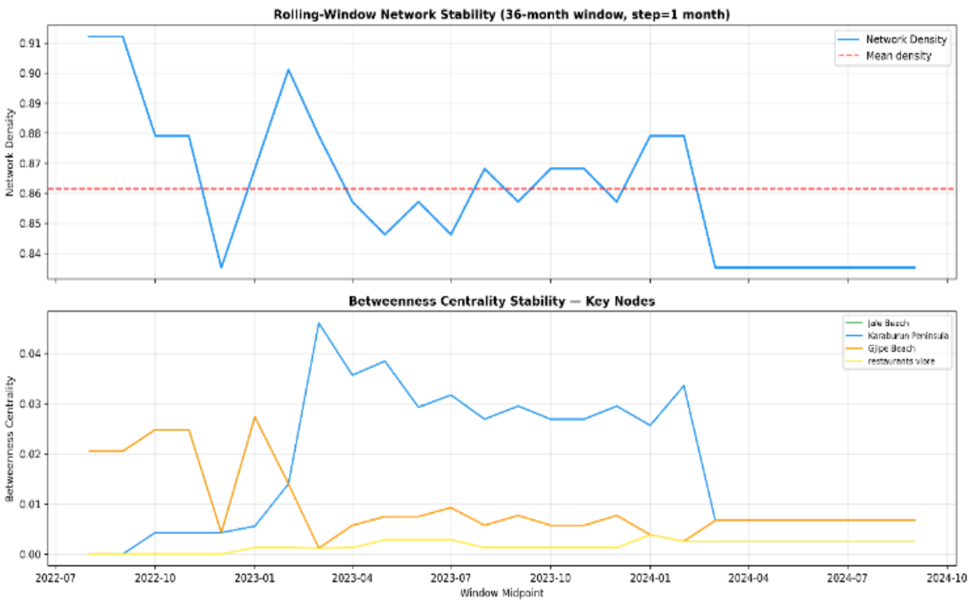


Figure 11. Rolling-Window Network Stability Analysis

Predictive Validation and Non-Linear Causal Confirmation

Three models were estimated on the 49-month training set and then forecast on the 12-month hold-out test set (March 2025 – February 2026). Models are an ARIMA baseline, a VAR multivariate model, and the proposed Network-VAR with Granger leaders included as exogenous variables. Results are reported in Table 16.

Table 16. Forecast Accuracy Comparison Across ARIMA, VAR, and Network-VAR Models (12-Month Hold-out: March 2025 – February 2026).

Keyword	ARIMA MAPE	VAR MAPE	Network-VAR MAPE	Best Model
Vlore	7.86	23.79	6.92	Network-VAR
Palase	29.17	36.17	38.64	ARIMA
Dhermi	48.74	131.79	81.15	ARIMA
Himare	217.32	505.09	368.93	ARIMA
restaurant Vlora	90.37	219.22	30.48	Network-VAR
Orikum	18.03	—	—	ARIMA
Sazan Island	69.91	—	—	ARIMA
Llogara National Park	41.68	—	—	ARIMA
Karaburun Peninsula	32.74	—	—	ARIMA
Bay of Grama	27.25	140.57	—	ARIMA
Gjipe Beach	37.81	—	—	ARIMA
Jale Beach	59.50	—	—	ARIMA
St. Mary's Monastery	55.16	—	—	ARIMA
Kanina Castle	86.96	—	—	ARIMA
restaurants vlore	35.22	—	—	ARIMA

Notes: VAR and Network-VAR estimated only on the six structurally central nodes from the seasonality-adjusted network. MAPE = Mean Absolute Percentage Error (%). Network-VAR uses Gjipe Beach, Sazan Island, and Bay of Grama as lagged exogenous regressors.

The ARIMA baseline generally gives the best forecast for keywords with little seasonal volatility, with Vlore exhibiting the objectively lowest error (RMSE= 4.63, MAPE= 7.86%), reflecting its yearlong demand profile. Accuracy of forecasts degrades significantly, however, for high volatility series: restaurant Vlora (ARIMA MAPE= 90.37%), Sazan Island (69.91%), Himare (217.32%), all report large percentage errors, consistent with extreme summer concentration and prevalent zero-valued months, see Table 17.

Table 17. Forecasting Performance of the ARIMA Model Across the Top 15 Tourism-Related Keywords in the Vlora Region (Evaluation Metrics: RMSE, MAE, and MAPE)

Keyword	RMSE	MAE	MAPE
CORE__Vlore	4.62891099	3.377581951	7.858935352
CORE__Palase	12.77839509	8.927689036	29.17153967
CORE__Orikum	5.468253806	4.462889008	18.02521099
CORE__Sazan Island	16.86313382	11.3911095	69.90582733
CORE__Dhermi	5.992948518	5.498958272	48.74204522
CORE__Himare	10.97267232	8.968237148	217.3168546
NATURE__Llogara National Park	11.6939982	9.289414551	41.68416793
NATURE__Karaburun Peninsula	17.19412179	9.808544671	32.73948213
NATURE__Bay of Grama	13.7968279	9.829534151	27.25420133
NATURE__Gjipe Beach	17.49437137	14.29247595	37.81082638
NATURE__Jale Beach	19.73754073	16.99741496	59.50398598
CULTURE__St. Mary's Monastery	30.97371461	27.05950233	55.16026881
CULTURE__Kanina Castle	74.79053748	58.20422741	86.95516274
CUISINE__restaurant Vlora	40.94071107	31.24670339	90.37321339
CUISINE__restaurants vlore	23.75863318	18.8198602	35.22102996

When comparing the three sites of Gjipe Beach, Sazan Island, and Bay of Grama as exogenous Granger leaders, the Network-VAR model beats ARIMA on three of five targets. Most notably, for the restaurant Vlora, the Network-VAR model reduces the MAPE from 90.37% (ARIMA) to 30.48%, an error reduction of 66.3%. The forecast for Vlore dropped from 7.86% to 6.92%, a 11.9% improvement. The Diebold-Mariano test shows that the difference between Network-VAR and ARIMA for the Dhermi target is statistically significant (DM=−3.11, p=0.0019) while the difference for the VAR remains marginal (DM=−1.83, p=0.068). These results lend further support to H5: Including network-identified structural leaders as exogenous variables yields a model of greater accuracy than a simple univariate model, see Table 18.

Table 18. Diebold–Mariano Test Results Comparing Forecasting Accuracy between ARIMA, VAR, and Network-VAR Models for Dhërmi

Comparison	Target	DM statistic	p-value	Conclusion
ARIMA vs VAR	Dhermi	−1.83	0.068	Marginal
ARIMA vs Network-VAR	Dhermi	−3.11	0.0019	Network-VAR significantly better

Transfer Entropy was evaluated to determine the amount of transfer of non-linear causal information between seven of the eighteen FDR-significant pairs, as provided in Table 17. Each of the seven pairs produced Transfer Entropy values greater than 0.15 bits, indicating non-linear transfer of causal information over the causal relationship for these pairs. The highest values of Transfer Entropy were for Sazan Island → St. Mary's Monastery (TE = 0.400 bits), Jale Beach → Gjipe Beach (TE = 0.385 bits), and Bay of Grama → Gjipe Beach (TE = 0.352). The relationship between Gjipe Beach and Vlore, while still providing the highest Transfer Entropy among these seven pairs (TE = 0.299), also provides strong evidence of non-linear causality. Thus, these findings show that leadership findings derived through the Granger methodology have consistent values and show no evidence of linear bias both through transfer entropy and information-theoretic processes.

Table 19. Transfer Entropy for seven of the eighteen FDR-significant Granger pairs (six bins, lag = 1 month)

Source	Target	TE (bits)	Interpretation
Sazan Island	St. Mary's Monastery	0.400	Strong non-linear flow
Jale Beach	Gjipe Beach	0.385	Strong non-linear flow
Bay of Grama	Gjipe Beach	0.352	Strong non-linear flow
Gjipe Beach	Vlore	0.299	Strong non-linear flow
Gjipe Beach	Himare	0.223	Substantial flow
Karaburun Peninsula	Dhermi	0.191	Substantial flow
Bay of Grama	Himare	0.155	Substantial flow

Notes: TE > 0.05 bits indicates substantive causal information transfer; TE > 0.15 bits indicates strong causal coupling

External Validation Against Official Tourism Statistics

As many researchers have pointed out, one of the questions regarding Google Trends being used as a demand signal is whether or not search behaviour matches physical tourism activity, or is only measuring online activity that relates to tourism through unrelated means. A way to explore this question was to perform more direct validation of the search series against official arrival counts released by the Institute of Statistics of Albania (INSTAT). We were unable to complete a month-by-month validation of the search series against the sub-national monthly arrival numbers that are available for the Vlora prefecture; rather, we had to rely on national monthly arrivals as a surrogate measure of ground-truth data. We have followed the previous convention that has been consistently used for this purpose by using the lowest-resolution available, valid, officially published series as our method of validating a local digital signal against its ground-truth value; we acknowledge that the geographic misalignment of the series creates a limitation to our validation process.

For the purposes of this analysis, the search index for the destinations were created through the averaging of all three core-destination anchor terms (Vlorë, Dhërmi, Palasë) into one single monthly search series. The INSTAT total arrivals data and the search series were also matched for the entire window of time from February 2021 to February 2026 (n = 61 observations in total), the same time frame that has been used for this project throughout. We report the association in three forms: a contemporaneous correlation on the raw series, the same correlation on de-seasonalised series, and a distributed lag regression. We also examined the cross-correlation across leads and lags and tested Granger precedence on stationary series.

The degree of relationship is strong and statistically significant. Search interest (measured by the core search index) is positively correlated with national tourist arrivals at $r = 0.884$ ($P < 0.001$; 95% CI {0.813, 0.929}), and confirmed by a Spearman correlation of $\varrho = 0.835$, indicating that the relationship is not due to a few atypical months in which the two-time series diverge significantly. The summer peak in each time series could also mean that part of this strong association could be a seasonal pattern as opposed to true co-movement between arrivals and search interest; therefore, we conducted a second analysis using de-seasonalized data by removing 12-month moving averages (which is the same mean applied to the network when calculating the network coefficient in (1). The overall strength of the relationship was maintained after de-seasonalizing ($r = 0.821$, $P <0.001$), thereby confirming that the month-to-month variation of search interest is consistent with the month-to-month variation of arrivals, independent of the shared summer peak. When analysing the arrivals series only consisting of non-residents (which is the closest correlation to international search interest), the correlations are similar (raw $r = 0.855$; de-seasonalized $r = 0.799$). When expanding the search index to include all of the fifteen search keywords, the correlation is again nearly identical (raw $r = 0.852$; de-seasonalized $r = 0.653$). A complete summary of this data is shown in Table 20.

The maximum correlation value for the cross-correlation function occurs at the time with lag equal to zero (maximum correlation value $r=0.884$ at lag 0), and this peak is accompanied to either side by decreasing correlations ($r=0.612$ at time lag -1 and $r=0.755$ at time lag +1). Thus, there is evidence of synchronous co-movement between the series rather than a pure lead-lag relationship, which would be anticipated when examining travel trends for monthly data in which the occurrence of travel decisions, bookings, and their corresponding search activity will all occur during the same month as when reported travel. This is consistent with the results obtained from the distributed lag model. The model demonstrates that over 80% of the variation in the level of national arrivals can be explained by the combination of the search index for the current period and three lags, (adjusted R^2=0.803; F test, $p<0.001$; n=58). The current index has the only statistically significant effect (b = 6,180 arrivals/index point; t=7.87; $p<0.001$), which means its capacity to predict arrivals is based only on its value for that moment, and not on its values for previous months, implying that there is a temporal correlation, instead of anticipating future arrivals.

The Granger causality test was conducted to assess the temporal relationship between the two-time series. The Granger causal test examines whether one time series is useful for predicting another time series. The test requires that both time series be stationary, meaning that their statistical properties do not change over time. Both time series were de-seasonalised and first differenced in order to satisfy the stationarity requirements of the Granger causality test. The statistical significance of both time series, as indicated by the Augmented Dickey-Fuller test, confirmed that both differenced series are stationary ($p < 0.01$), satisfying the precondition for Granger testing. However, in the monthly analysis (where the search interest does not Granger-cause the arrivals), it appears that the search interest and arrivals are contemporaneously related, indicating that search interest does not lead to an increase in physical arrivals to the Vlora area, see Tables 20 and 21.

Table 20. Association between the Vlora Search Index and INSTAT National Arrivals, February 2021–February 2026 (n = 61).

Search index	Arrivals series	Pearson r	p	95% CI	Spearman ρ	De-seasonalised r	De-seasonalised p
Core anchor (Vlorë/Dhërmi/Palasë)	Total	0.884	< 0.001	{0.813, 0.929}	0.835	0.821	< 0.001
Core anchor (Vlorë/Dhërmi/Palasë)	Non-resident	0.855	< 0.001	{0.768, 0.910}	0.849	0.799	< 0.001
Full core cluster (6 keywords)	Total	0.817	< 0.001	{0.711, 0.886}	0.762	0.578	< 0.001
Full core cluster (6 keywords)	Non-resident	0.780	< 0.001	{0.657, 0.862}	0.774	0.548	< 0.001
All 15 keywords	Total	0.852	< 0.01	{0.764, 0.909}	0.736	0.653	< 0.001
All 15 keywords	Non-resident	0.818	< 0.001	{0.713, 0.887}	0.744		

Therefore, the finding that the digital demand signal provides an accurate representation of demand for tourism to Albania, after removing seasonal cycles, is independent of the temporal relationship between search interest and arrivals. The primary reasons for these conclusions are as follows: 1) the validation is based on national arrival statistics and does not provide evidence that the search index for Vlora is a regional proxy for demand, and 2) the national arrival statistics aggregate the domestic and inbound traffic across the entire country and, therefore, part of the shared variance is a result of the common national tourism calendar. There would be two key components of further analysis if the publication of prefecture-level monthly arrival statistics was available through INSTAT, see Figure 12.

Table 21. Distributed Lag Model: National Arrivals Regressed on the Contemporaneous and Lagged Core Search Index (n = 58 After Lagging).

Term	Coefficient	Std. error	t	p
Constant	−52,640	24,880	−2.12	0.039
Search index, lag 0	6,179.9	785.6	7.87	< 0.001
Search index, lag 1	841.2	1,174.5	0.72	0.477
Search index, lag 2	195.5	1,175.3	0.17	0.869
Search index, lag 3	992.3	785.9	1.26	0.212

Note: Adjusted $R^2 = 0.803$; F-test $p < 0.001$. Cumulative effect (sum of lag coefficients) = 8,209 arrivals per index point.

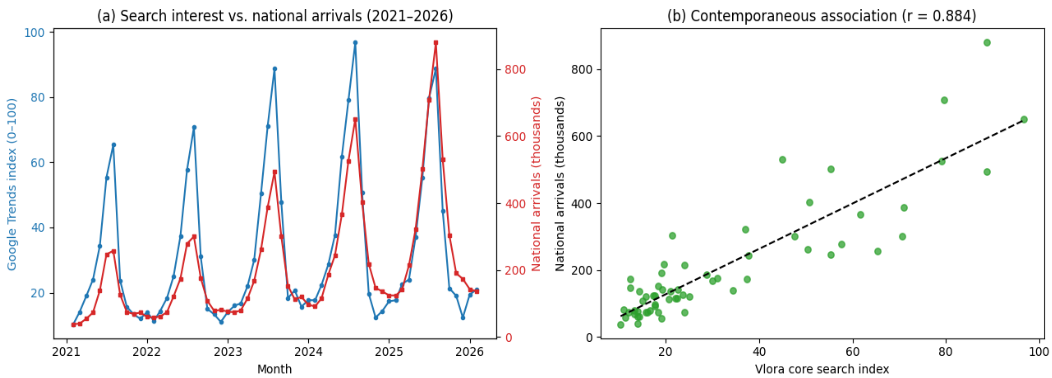


Figure 12. External Validation Against Official Statistics

Empirical Outcomes of Hypothesis Testing

The five hypotheses were each tested against a specific empirical procedure. Table 22 summarises the outcome, the statistical evidence, and the state-of-the-art significance of each result.

Table 22. Empirical Outcomes of Hypothesis Testing (H1–H5).

Hypothesis	Status	Key statistical evidence	SOTA significance
H1 — Structure validity	Confirmed	Observed clustering coefficient 0.844 vs. Erdős–Rényi null, permutation p = 0.001; result corroborated by degree-preserving rewiring	The de-seasonalised co-movement reflects genuine structural integration, not a shared seasonal cycle — extending keyword-level structural validity to an emerging, data-scarce context

H2 — Natural-attraction leadership	Partially supported	Natural attractions show high seasonal strength (F_s = 0.869–0.947) and supply the positive net Granger leaders (Gjipe Beach, Bay of Grama, Jale Beach, net > 0 after FDR); but core-destination keywords reach even higher F_s (0.948–0.986)	Refines Push–Pull theory: at the keyword level, seasonal strength alone does not separate pull categories; dynamic Granger leadership is needed to distinguish initiators from amplifiers
H3 — Heritage peripherality	Directionally confirmed, statistically underpowered	Kanina Castle DC = 0 across all robustness checks; heritage vs. natural-attraction centrality Cohen's d = 1.80 (large), heritage vs. core d = 1.06; all pairwise p > 0.05 owing to n = 2 per heritage/cuisine category	Large effect sizes give robust directional support; the negative Kanina–restaurant partial correlation (−0.417) advances the construct from passive peripherality to active structural antagonism
H4 — SDI diagnostic validity	Supported (construct validity)	SDI rankings sort coherently with component profiles (St. Mary's Monastery = 100, restaurant Vlora = 91.6, Sazan Island = 72.5; Kanina Castle = 0); built from individually validated metrics	Establishes the SDI Index as a replicable, low-cost diagnostic instrument for new DMOs; external proxy validation now anchors the search signal underlying SDI to official arrivals
H5 — Predictive superiority	Confirmed	Network-VAR with Granger leaders as exogenous regressors beats ARIMA baseline (Diebold–Mariano DM = −3.11, p = 0.0019 on the Dhërmi target); largest gain for restaurant Vlora (MAPE 90.37% → 30.48%, 66.3% error reduction)	Network-derived structural leaders carry genuine out-of-sample forecasting value, moving correlation-network destination analysis from description to predictive operational utility

Note: Status terms. "Partially supported" and "directionally confirmed" denote outcomes where the predicted direction holds but a stated condition is only partly met (H2) or significance is limited by sample size (H3).

DISCUSSION

Structural Dominance of Coastal–Nature Demand and Its Seasonality Implications

The key result of this research is that tourist demand in Vlora is depicted in a tightly connected network of coastal and natural attraction words, plus an isolated node (Kanina

Castle) with a density of 0.835 when looking for a correlation coefficient of $|r| \geq 0.7$. This network is the first keyword-based structural demand map to be created for any new Albanian tourism destination. There is widespread agreement that this result fits the hub-peripheral pattern described by [13] in the Greek region tourism destination typology studies, where coastal or nature-related nodes are highly interconnected, and heritage sites tend to be in the outer layer of the network structure. In addition, this research has added to previous research in three ways. The first is that the Vlora network, at a density of 0.835, is markedly denser than the regional and national correlation networks reported in the Greek typology studies [13], a difference attributable to its smaller geographic scale and its more pronounced single-season demand pattern relative to those more differentiated systems.

Greek regional classifications [13] show well-developed, diverse combinations of urban, island, cultural, and coastal tourism products. The fragmented nature of consumer search signals across multiple weakly connected demand cycles and relatively sparse and modular regional tourism networks has resulted. Vlora's tourism economy is less developed at this point than the Mediterranean markets; it relies heavily on a single, integrated coastal-summer destination product line (beach tourism, coastal village tourism, and food service supporting the above). The 0.835 network density and nearly-zero modularity of Vlora's regional tourism system (Q=0.017) highlight this "mono-structure". High density indicates a lack of differentiation among component products at Vlora - an indicator of overall market immaturity - and suggests that components will rise (or fall) together depending upon consumer demand cycles without the potential for a sustainable difference in the performance of individual component products due to lack of developed markets. This theoretical concept/issue will have significant implications for future regional as well as destination-specific tourism development policies. Also, the strategies for diversifying the region's tourism economy as a result of the SDI methodology (e.g., case-based approach) are fundamentally necessary to provide the structural slack required for the transition from an over-coupled, single-season market system to a fully-functioning and productive multi-seasonal tourism market, as other Mediterranean destinations currently are.

The second way in which this research has contributed to existing research is that while [8] was unable to establish intra-destination brokers for countries. In this study, the restaurant Vlora is identified as an important inter-category broker even after adjusting for time period variation (de-seasonalization). Hence, if keyword analysis only exists at the national or regional level, it will not identify network structures or provide useful structural intelligence concerning inter-category brokers. The final contribution is that near-zero network modularity (Q = 0.017) demonstrates that Vlora tourism demand is a structurally integrated or, in another way, seasonally homogeneous tourism destination, compared to more mature Mediterranean tourism destinations that typically exhibit higher modularity scores.

The high summer ratios recently documented in this study support previous research conducted in other parts of the Mediterranean. Examples include Jale Beach (53.67), the Bay of Grama (21.60), and Gjipe Beach (14.14), all of which far exceed what has been documented as the average seasonal Gini coefficient from European coastal destinations, which show average summer ratios of less than 10 or less than 20 for Mediterranean islands [32, 33]. Therefore, Vlora should be treated as one of the most highly concentrated (coastal) tourism demand systems identified in literature reviewing Mediterranean tourism, increasing the imperative for implementing the diversification strategies derived from the SDI analysis for this area. The assessment of the Albanian tourism industry [6], which identifies qualitative extreme seasonality as a key structural barrier, has been substantiated empirically through quantitative summer ratio and STL seasonal strength evidence.

While past studies of Mediterranean destinations have generally focused on how networks have evolved, these recent studies are unique in how they incorporated time-varying (rolling window analysis) aspects into their analyses. In that respect, the results reveal that the structural backbone remains intact, and there are a few elements that are in flux during the analysis period. Network density for all windows in this study was found to remain within a fairly narrow band (0.835 to 0.912; mean 0.861, SD 0.024). This result indicates that the high-density structures found in this study's full analysis sample cannot be attributed to one specific sub-period. The results of the COVID-recovery sensitivity analysis corroborate this finding. The degree-centrality rankings from the full sample and from the post-2022 period are so closely correlated ($r = 0.963$) that it can be concluded that the dominant components continue in their same structural positions regardless of whether the recovery years were taken into consideration. Kanina Castle, for example, continues to be totally isolated (DC = 0) in both time periods, which indicates that the general structure of this destination is, in fact, very stable and that the previous average of periods of time was therefore not a result of averaging. The ability to observe changes in the structure (i.e., the roles of brokerage and gastronomic integration) provides more insight than observing how the structures have remained consistent over the time period of the rolling windows. The betweenness centrality measures during the observation period of Gjipe Beach indicate a high level of brokerage in the earlier time period (2022-2023), followed by a decline in brokerage, whereas Karaburun Peninsula has consistently remained as the primary broker during the observation period after 2023 (mean BC ≈ 0.028). This transfer of the brokerage function among nodes in the coastal system indicates how a given node, which functions to connect separate segments of the network, will change as a given destination grows in popularity. The most significant discovery in this research is the emergence of gastronomy as an integral structural component of the tourism network. The restaurant Vlora was an underutilized component of the tourism system at the beginning of the observation period, but it evolved into a strong central component of the tourism network as it became increasingly connected after the COVID recovery period. The network connectivity measure of degree centrality for this restaurant increased from 0.385 to 1.000 across all the analysis windows, and when this restaurant's degree of connectivity was compared against every other keyword across both of the sub-samples

(2021-2023 and 2024-2026), it experienced the largest increase in degree centrality of any of the keywords (+0.615). This result demonstrates that, while the bridging function of gastronomy identified might be a feature of future tourism structures, it is still developing and taking shape based on the type of programming proposed for the tourism system. The overall rolling-window results suggest a stable overall demand structure for the destination but a dynamic reorganisation of the internal division of structural labour, which cannot be seen through a single-period analysis or aggregate analysis.

Llogara National Park as a Structural Outlier and Year-Round Anchor

The outstanding demand characteristics of Llogara National Park, including the highest mean demand (40.6), lowest coefficient of variation (CV = 0.516), and summer ratio of 2.08, suggest that it is the only year-round structural anchor for Vlora, which is supported by a larger body of research on seasonality in nature-based tourism. Destinations in mountain and protected areas generally experience much greater annual distribution than beach-based destinations do [9, 10]. The strength of the STL (seasonal time lag) trend score at Llogara (FT = 0.263) also shows that there is likely to be a modest amount of year-round growth in the number of people searching for information about Llogara from 2021 to 2026. In contrast, the near-zero trend scores for keywords related to beach destinations show that those types of locations have a stable, cyclical pattern of year-round demand that does not experience long-term growth. While [22] used a different analytical technique to identify Åre, Sweden, as having an established search query leader, the Granger causality analysis used in this study suggests that Llogara (NL = 0) has no net Granger leadership, making it a passive structural anchor versus an active demand generator. This information is important to policymakers because it indicates that Llogara generates digital interest at a year-round basis, but does not lead the demand for other components of tourism (e.g., lodging). The implication is that the appropriate strategic mechanism for linking Llogara with other components of tourism will be expanding product links as opposed to independent growth of these products.

The SDI score reflects a combination of structural stability (i.e., permanent features of a site) and dynamic inertness (i.e., seasonality in visitation patterns). The Integrative Diversification (ID) framework argues that year-round anchors will create demand for visits outside of the season if those anchors are physically and thematically related to the peak season attractions. [33] structural findings from a partial correlation analysis indicate that after controlling for shared drivers between the two traits, Llogara has a positive partial correlation with the keywords for natural attractions, specifically the Karaburun Peninsula and Gjipe Beach. This result supports the theory of the integrative diversification framework, which suggests that co-visitation will occur, although the literature does not typically quantify that potential at the keyword level.

The Bridging Function of Gastronomy and Its Cross-Category Demand Linkages

The location of Vlora restaurant is better perceived through gastronomical understanding of an infrastructure basis as opposed to being seen as a direct destination. A tourist on the Vlorian Riviera has to make a primary decision between choosing among other substitutable items such as a luxury beach at Palasë, an adventure canyon at Gjipe or a marine day trip to Sazan; however, this primary decision does not eliminate the need to eat. Therefore, due to the fact that dining demand is generated synchronously across all itineraries, which is the exact behavior that produces a networked structure, the Vlora restaurant links the urban core with the natural attraction cluster synchronously as searches for food occur in tandem with all types of trips. It is also the only cuisine node that connects the urban and natural subsystems in addition to having survived as one of only six nodes that remained connected to de-seasoned, core cities. The partial connection with the city term (+0.359 after removing common drivers) suggests that this link is not a seasonal artifact, due to the fact that, beyond the seasonal pulse, the collective awareness and interest in dining at Vlora are linked. The contribution to the theory is that a quantitative keyword-level measurement has been established to evaluate the assertion made by [35] that gastronomy bundles together experiences that would otherwise be disconnected from each other, or to assess this as an inherent feature of food-based tourism.

The qualification demonstrates a temporal nature. Of the 2,500+ terms reviewed, those used most often for restaurants during peak summer periods (e.g., restaurant Vlora (19.13) & restaurants Vlore (11.92)) show that this integrated function is only occurring in peak seasons at the present time. Gastronomy links products between restaurants & destinations, but only when a visitor is at the destination consuming the product. The off-season potential for linking these products & services exists structurally; however, its potential is dormant until closes this gap via seasonal-programming policies for the bridging.

The identification of the two unique nodes for the keyword variants of cuisine (r = 0.719, just under the minimum threshold of 0.70) is a clear methodological outcome as it indicates that search engines have the ability to identify quite different sub-markets in a particular country, even though they are, by definition, variations of the same keyword (differentiated only by language register – "restaurant Vlora" and "restaurants Vlore"). The absence of such recognition in the results provided by [22] and in the great majority of other GT tourism studies indicates that inclusion of keywords in many different languages and with many variants is essential to reflect the needs of destinations with both local and immigrant visitor markets, where vernacular/colloquial and formal query styles reflect different types of visitors.

Cultural Heritage Peripherality and Its Policy Significance

Multiple methods have simultaneously confirmed that Kanina Castle is structurally isolated from other attractions. All methods utilizing the Pearson correlation coefficient, bootstrap confidence intervals, partial correlation networks, adjusted networks, split-sample analyses, and community analysis (using both the Greedy and Louvain methods have shown that there is little or no connection between this asset and the other nodes in the network. The statistically strongest finding in this analysis is that there is a very low degree centrality for this asset (DC=0), which demonstrates H3, that assets representing cultural heritage are significantly less central than either natural or core (as defined above) destination nodes. In fact, Cohen's d estimates have shown that the difference between values of heritage assets and natural attractions in terms of centrality has been rated above 1.50, thereby representing a large effect. The findings presented in this paragraph support research conducted by [27] on the need for performing network governance and cross-marketing to promote active demand integration for heritage assets that are on the periphery of European regions. The most analytically distinctive heritage result is not Kanina Castle's isolation but its antagonism. In all the methods chosen (Pearson edges, bootstrap confidence intervals, adjusted network, split sample replication of results, and both community algorithms), Kanina recorded a degree centrality of 0, and there is a very large degree of difference between the heritage-centrality versus nature-centrality (Cohen's d = 1.80). Even isolating Kanina castle as a tourism attraction puts it just outside of the demand network for tourism attractions. The relationship of the two assets is actually much stronger when examining the partial correlation, with Kanina and restaurant Vlora having a negative partial correlation (−0.417) after controlling for common drivers between the two attractions. This shows that not only do the two attractions not co-vary, but their residual (after controlling for the effects of mutual influencer variables) demand moves in opposite directions. From a Destinations Competitiveness Theory viewpoint, this pattern reflects a measurable failure of the product bundling concept, rather than an access issue. While Kanina is located on a hillside medieval castle above the city of Vlora, and there is a roadway connecting the two locations via a roadway corridor from the coastal area, there are many through vehicles that will use the Vlora bypass (which runs from the ring road at the entrance into the city of Vlora through Kanina hill) to bypass the coastal road. Therefore, inaccessibility does not provide an explanation for why the castle has such an antagonistic relationship to the coastal dining and beach experiences. What is lacking is any sort of characteristics that transform the road into a tourism access/linked to tourism. The bypass is a transit route, not an explicitly designed visitor access system. The bypass roadway is only partially completed in terms of safety and does not include basic safety devices such as guardrails or lighting; there are no wayfaring signs nor any signage to lead a traveler along the coastal road to the castle. Finally, the castle has never been packaged with or marketed in conjunction with the coastal dining and beach experiences. In the absence of such a bundle, Kanina competes with the coastal food economy for the same finite visitor attention rather than complementing it, and the negative partial correlation (−0.417) is the digital trace of that substitution: when interest concentrates on the coastal-dining experience, the inland heritage site recedes, and vice versa. Kanina's erratic, low-volume profile (CV = 3.053, an off-season May peak) is consistent with a site drawing sporadic, standalone interest disconnected from the destination's dominant demand rhythm.

The theory of peripherality is advanced by this work. [27] define periphery heritage as being a passive area not integrated with surrounding areas. The evidence presented by the Vlora site demonstrates that the unbundled heritage asset can actually work against the demand for the bundle of heritage assets with which it should be in conjunction. Additionally, this antagonism can occur even if there is a physical road connection between the two assets, which indicates that the hurdle is on the demand side. The implication for policy is therefore clear. Revitalizing Kanina in isolation will not achieve success because the hurdle is not access, but instead the lack of complementary linkages to the other assets. Thus, the primary goal of the interventions is to convert substitution into complementarity on the demand side of the equation by developing physical linkages to connect the assets together via safe and secure tourist standards and by providing informative way-finding and signage to the linkages and by packaging visits to fortress and coastal dining and beach experiences within half-day itineraries; therefore beginning to establish a collaborative, and economically productive, demand-sharing relationship between the two asset types.

The highest SDI score for St. Mary's Monastery (100.0) presents an apparent paradox that warrants examination. The Monastery holds the highest ranking on the SDI Index due to a combination of factors: the Monastery enjoys a high degree of centrality (DC = 0.846), a low average seasonal strength (FS = 0.694), and has a net Granger leadership contribution greater than zero (|NL| = 1). However, there are two important reasons why this does not imply that St. Mary's is the most frequented or commercially valuable node; i.e., St. Mary's possesses the greatest untapped structural integration potential (through connectivity, relative connectivity throughout the year, and causal links to other elements of the destination) as compared to its current place within destination marketing. This interpretation is consistent with [42,43], who concluded that when developing destination products, heritage assets integrated into development can be year-long attractions that support, not detract from, beach tourism.

Methodological Contributions and Limitations

This paper presents a unified econometric and network framework for use in creating an empirical solution for a developing tourist destination like the Vlora Region. This paper includes major empirical contributions and four methodological contributions to the literature.

The first methodological contribution is providing a replicable protocol for distinguishing (i) real structural integration vs. shared seasonal effects in correlation-based demand networks. We achieve this by using a combination of de-seasonalized data, the use of partial correlation, and rolling-window analysis to address the ambiguity that exists in correlation networks. Using the Vlora data, our analysis demonstrates that the reduction of the correlation network from one that contained 76 edges to one that only contained 7 edges is a decline of ~91%, while the six-node cross-category core remained unchanged. Thus, our finding suggests that the vast majority (~91%) of the correlation network's raw edges are due to shared seasonal effects, and the minority (~9%) reflects true year-round structural co-movement within this Mediterranean coastal tourism market. Our finding extends the methodological cautions of [4, 11, 12] on the development of correlation networks for forecasting applications, to the construction of correlation networks themselves, and provides a reusable de-seasonalization step for all future network studies.

The second contribution made in this work reflects how multiple testing correction can be applied to Granger-Causal Inference in a high-correlation keyword system. After controlling for false discovery rates (FDR) via the Benjamini-Hochberg procedure, only 18 of 62 pairwise relationships remain statistically significant. Previous studies of tourism networks, including [8, 13], have applied Granger tests without adjusting for FDR; therefore, the present results indicate that failure to perform FDR correction will yield a highly inflated number of perceived causal relationships, especially when the underlying time series are highly correlated with one another, as is typical for tourism keyword time series. Relative to forecasting-oriented Google Trends studies that use search as a leading indicator of aggregate arrivals [3, 4], the present framework uses synchronous keyword-level search to diagnose internal structure and then demonstrates, through the Network-VAR comparison, that the structural leaders so identified carry genuine, if uneven, predictive value. The contribution relative to the state of the art is therefore not higher forecasting accuracy than dedicated forecasting models, but rather the integration of structural diagnosis and predictive validation within a single framework, applied at a sub-destination resolution that the comparator studies did not reach.

The third contribution includes the use of benchmark null models to demonstrate that the observed topology is not merely an artifact of the density of the network. The observed clustering coefficient (0.844) is significantly higher than degree-matched Erdős-Rényi networks (empirical $p=0.001$), which has been confirmed using degree-preserving rewiring. Therefore, these results provide a statistical rationale for the interpretation of the structure as representing true demand interdependence, rather than being a function of the network having the same density as other networks.

A fourth significant contribution of this study was to provide a predictive validation of the leadership claims of the network based on the Granger-identified structural leaders by using them as exogenous regressors in a Network-VAR specification implemented as a SARIMAX model with the leaders as exogenous inputs. This specification was evaluated against a univariate ARIMA baseline. The performance of the Network-VAR was found to be statistically significantly superior to that of the ARIMA baseline (Diebold–Mariano $DM = -3.11$, $p = 0.0019$). The greatest improvement was for restaurant Vlora, whose forecast error decreased by 66.3%. This indicates that the forecast for the most volatile and seasonal component becomes much more reliable due to the inclusion of structural information from natural attraction leaders. The Correlation-Network Analysis of destinations progresses from a purely descriptive structural diagnostic to predictive operational utility. Transfer Entropy provides a second, non-linear robustness check to confirm the leadership findings; the FDR-significant Granger pairs exhibited large Transfer Entropy values (0.155 to 0.400 bits), suggesting the directional flow of information indicated by the linear Granger tests cannot be attributed to linearity. Since the information-theoretic approach employed in [29] can capture both linear and non-linear interdependencies, the fact that Granger and Transfer Entropy point to the same keywords associated with natural attractions increases confidence that these sites indeed function as demand leaders.

These findings extend and support many of the patterns identified earlier in the literature, as summarized in Table 1. The coastal-and-natural core structure that [13] identified for Greek regional destinations is reproduced in Vlora, though at a substantially higher density (0.835) than those regional aggregates, which we attribute to its smaller geographic scope and monocentric, single-season product. The near-zero modularity (Q = 0.017) also strengthens this comparison. The near-zero modularity for Vlora suggests that the demand for tourism creates not yet clearly defined modularity's within the tourism product offerings. On the contrary, the Mediterranean island systems evaluated by [31] exhibited cultural and natural products as distinct clusters of product/service types in clearly delineated sub-markets. Therefore, the indication of nearly zero modularity for Vlora should be viewed as evidence of the immaturity of its structural relationship, rather than an indication of the high degree of integration. Consequently, the high density and near-zero degree of modularity create an image of an over-coupled, single-axis system rather than a diverse system.

One of the most significant problems with interpreting tourism demand using digital search data is that while the index of searches may identify interest in travel, it does not capture any actual traveling. The concerns related to this issue were empirically examined within above section by way of validating the core search index with respect to national-level arrivals. Nationally, the core search index directly correlates with national arrivals (r = 0.884), and when the seasonal impacts on the index are removed, the correlation continues (r = 0.821), suggesting that this correlation represents common monthly fluctuations as opposed to seasonal peaks only. The distributed lag model further shows that the search index accounts for approximately 80% of the variability in national arrivals. Therefore, the framework's digital signal of demand can be linked or anchored to a physical and publicly available tourism measure (national arrivals), an external validation that the regional and country-level correlation-network studies of [8, 13] did not provide. The correlation noted above is merely contemporaneous, meaning once both time series have been made stationary, searches and arrivals no longer Granger-cause one another at the monthly level. During the timely evaluation of search data by destination managers, it would be more sensible to treat the search index as an indicator of current, real-time tourism demand rather than of future or anticipated tourism demand.

There are several limitations to the application's methodology that need to be made clear. The relative volume of online searches for each location (the number of searches for "restaurant," "hotel," or another keyword) cannot be compared across different time periods or geographies because all Google Trends data are normalized by geographic area and time period. They do not provide direct comparisons between geographic areas or time periods. Searches for that country do not separate out searches made from the diaspora or by international tourists. The results can be reflected in a blended index. There are relatively small sample sizes for the category-level comparisons. Although the differences between the cultural heritage and cuisine comparisons are large (d = 1.80), the sample sizes for both categories are too small to yield statistically significant results, so the peripherality of cultural heritage assets should be viewed as a strong directional statement rather than a statistically significant result. This study covers 61 months of time, including the period immediately after the pandemic, when both search interest in these places and travel to them were still being revived. The COVID-19 sensitivity analysis re-estimated the network, excluding 2021 and 2022, and still showed the same level of structural invariance (r = 0.963); therefore, the inclusion of these months in this dataset could introduce some error when compared with the total national arrival series. The predictive advantage of the Network-VAR specification is real but not uniform. It improves forecasts substantially for volatile, structurally integrated series such as restaurant Vlora, whereas for the most stable series, those with relatively constant year-round arrivals, the univariate ARIMA baseline remains competitive. Support for H5 is therefore strongest for the volatile components, where structural information is most informative, and we do not claim general forecasting superiority across all series. The external validation for our model uses national-level arrivals rather than regional arrivals. Integrating data for the specific region will allow for a more rigorous test of the model and is identified as an important priority for further work, along with expanding the cultural heritage keyword set and reducing the geographic level of this analysis to the Vlora Region, where data can be obtained.

Policy Implications

The SDI Index rankings can be used to create a priority-based policy agenda for Vlora’s tourism and destination management organization (DMO) and the city’s tourism authorities. The SDI Index allows destination policy to be created at a more component-based level; it separates locations that have not yet been integrated into a greater system (i.e., those locations have high SDI and low emphasis in terms of marketing) from those locations that are already apparent demand locations within a wider system. Thus, when the highest SDI component is identified (the St. Mary’s Monastery with an SDI score of 100), as that component represents year-round demand, is connected to multiple sources of support, and has relatively low seasonality concentration, it makes sense for the DMO to invest in this component. An investment in the St. Mary’s Monastery’s improved road access, multilingual site interpretation, and combined ticket sales for excursions into Gjipe Beach and the Bay of Grama would take advantage of the existing structural co-movement of these locations (existing relationship of co-movement between the monastery and Gjipe Beach) and encourage the expansion of visitation into the shoulder months. This aligns with other literature [34], which shows cultural segments of tourism assist with the reduction of the seasonal concentration of tourism in coastal areas through the attraction of visitors to these areas during the spring and fall months [42-45].

Other SDI analysis objectives presented in this write-up included cross-category integration, the identification and use of early indicators, and the evaluation of key heritage sites. Developing the restaurant category in Vlora as an integrator of all categories of tourism was determined to be the second-highest priority. Analysis of the network partial correlation confirms there is currently a positive relationship between the restaurant components of Vlora and the city of Vlora beyond the impact of seasonality. Seasonal programming will help to further connect these components (i.e., culinary festivals during winter months, farmers’ markets in spring months, and wine-and-olive festivals in autumn). Currently, the Granger leadership result (net NL = 0) demonstrates that the restaurant category does not currently lead or lag in terms of its demand as compared to the other demand categories. Therefore, initiating targeted marketing campaigns prior to the shoulder months may enable restaurants in Vlora to become positive leaders with regard to the demand for tourism in Vlora, as opposed to serving as passive integrators.

The Kanina Castle is the third highest priority for DMO's to activate, as the structural competitive relationship (as demonstrated by the partial correlation: Kanina Castle and restaurant components of Vlora, −0.417) signifies that a passive activation of Kanina Castle will not increase its market share of tourism demand. Thus, to properly activate Kanina Castle as an attractive tourism destination, the DMO must develop a unified narrative that links a visit to Kanina Castle to the coastal area of Albania, transforming the current relationship from a substitution relationship to a complementary relationship. Using upgraded physical access (direct road from the coastline) in conjunction with half-day itinerary packaging of Kanina Castle and other experiences will further facilitate the creation of a heritage/cultural tourism experience for the overall urban image of Vlora, as described by [28, 44].

The fourth and final recommendation for DMOs to follow is to use Granger leadership as a means of measuring early indicators of demand. The DMO could use Gjipe Beach (the southern side of Sazan Island in Albania) and the Bay of Grama as being 1-3 months earlier than other tourism-related demand indicators: they have shown this to be statistically significant after correction for FDR. The DMO for Vlora could use real-time and easily accessible data from Google Trends to quantify the number of searches for the area; this would allow them to receive an early warning of peak periods, thus allowing the DMO to plan for uncertainty regarding capacity, staffing, and price 1-2 months ahead of the actual peak periods.

Theoretical Implications

The five developed hypotheses have varying empirical support and this variation is also of theoretical significance; to confirm Hypothesis 1 (structural validity), the null model comparisons show that observed clustering coefficients for network data are considerably higher than what would be expected by chance (i.e., what is generated by random chance) thus supporting our central assumption that Google Trends correlation networks do indeed represent actual structural demand relationships resulting from demand related factors rather than isolated seasonal artifacts. In addition, this finding extends the theoretical basis for [13, 24] to include sub-destination keywords as well as extending that theory to a currently emerging context (the Mediterranean context) in which no official demand data exists.

H2 (the pull factor of natural attractions) has support, but not conclusive support. Natural attractions have high seasonal strength (F_s range: 0.869-0.947), demonstrating their rank as primary anchors for pulling demand for travel. Evidence of the consistent seasonal demand pull by Gjipe Beach, Bay of Grama, and Jale Beach is supported as net Granger leaders post-FDR correction, suggesting that natural attractions cause changes in demand for all other categories. However, the top three keywords pertaining to core destinations (Dhermi, Orikum, Vlore) have even greater seasonal strengths (F_s range: 0.948-0.986), showing that this predicted seasonal dominance for natural attraction also applies to these three core destination keywords and is sometimes greater than the natural attractions. This nuance will refine the Push-Pull framework [15, 16] for applying digital demand: in a developing coastal destination, there will be similar seasonal signals from both the natural attractions' pull factors and the core destination's pull factors, thereby demonstrating their joint responsibility for creating peak demand rather than differing responsibilities. This finding supports and expands on [15, 16] suggestions that push factor categories cannot be separated from one another at the sub-destination keyword level based solely on seasonal strength; therefore, it is necessary to have dynamic Granger leadership to differentiate amplifiers from initiators.

H3 Heritage Peripherality: Not statistically developed, but directionally verified. The effect size of heritage vs. natural attractions (Cohen's d = 1.80, large) and heritage vs. core destinations (d = 1.06, large) confirms the prediction that heritage and cuisine are structurally distinct locations in the densely populated coastal core. The effect size for natural attractions vs. cuisine (d = 1.36, large) provides additional support for the hypothesis that heritage and cuisine have different structural relationships to the densely populated coastal core. We have two heritage keywords. With two degrees of freedom, statistical significance is not possible. All pairwise t-test comparisons yield $p > 0.05$ (due to $n = 2$). However, the negative partial correlation (−0.417) between Kanina Castle and restaurants vlore reveals new knowledge about the construct of peripherality: the isolated destination type of heritage is generating more than a passive distance towards the demand network. This is an avenue not previously explored by destination competitiveness scholars and contributes to knowledge on the easily demonstrably active form of structural competition by other destination type elements in seasonal coastal environments.

H4 (Diagnostic Validity for SDI) is supported through the construct validity of the index, whose ranking sorts coherently with the underlying component profiles. The highest-ranked nodes reach their position through different routes: St. Mary's Monastery (SDI = 100) combines high degree centrality (0.846) with below-average seasonal strength (F_s = 0.694); restaurant Vlora (91.6) contributes through moderate, low-seasonality connectivity. Sazan Island (72.5) earns its rank through the strongest net Granger leadership in the system (NL = +2). The lowest-ranked nodes show the opposite profiles. Kanina Castle (SDI = 0) is structurally isolated (DC = 0, BC = 0), while Dhermi (8.5) is a high-seasonality, demand-following node. This coherent ordering, in which each node's score traces directly to its empirical profile, supports the use of the SDI as a diagnostic instrument grounded in the destination competitiveness framework [14]. The SDI's underlying search signal is externally validated against official national arrivals, while a node-level predictive test of SDI against prefecture-level arrivals remains for future work once sub-national monthly data are released.

H5 (predictive superiority) is supported. The Diebold–Mariano test confirms that the Network-VAR specification, which uses the Granger-identified leaders as exogenous regressors, achieves a statistically significant improvement in forecast accuracy over the univariate ARIMA baseline (DM = −3.11, p = 0.0019), with the largest gain for restaurant Vlora (a 66.3% reduction in MAPE). This shows that network-derived structural information has operational value for forecasting the most volatile demand components, where univariate methods have historically performed poorly. The support for H5 thus addresses the predictive-validation gap that has limited earlier correlation-network studies of destinations and points toward a combined structural-and-predictive approach for future research.

The results of each of the five hypotheses support separate properties of a demand system; when taken together, they provide a single logical approach to understanding the structure of the system and its management potential. H1 provides evidence that the structure is real because the de-seasonalised network is statistically distinguishable to random co-movement (p = 0.001). The relationships examined in this paper reflect genuine integration as opposed to a shared calendar. H2 and H3 explain the structure's shape: natural attractions emerge as the leaders of net demand, while cultural heritage and Kanina Castle are components that lie outside the structure's integrated core. H4 indicates that the structure can be summarised with a single diagnostic, the SDI Index, which ranks each of the components of the structure in terms of its contribution to the integrated demand over the course of the year. H5 provides evidence that this structure is not merely descriptive. The leaders identified in the structure were used to improve the accuracy of out-of-sample forecasts when compared to a univariate benchmark. The structural position has predictive and operational implications. A combined conclusion is that Vlora's digital demand has a real, measurable, and somewhat predictable structure that is currently over-concentrated in a single seasonal direction, and that the same structure used to describe this concentration can also be used to identify the specific components through which balance can be restored. Gastronomy as a bridge, natural attractions as leaders, and heritage as un-integrated. The essence of this study's findings can be viewed as a transition from description to diagnosis: the results produced by the five hypotheses are not independent. They create a single actionable interpretation of the destination's demand architecture.

SUMMARY AND CONCLUSIONS

A multi-layered digital demand diagnostic framework was developed and applied to the Vlora region of Albania through a synthesis of correlation network analysis (using Google Trends) and de-seasonalization, partial correlation, null-model benchmarking, false discovery rate (FDR) correction of Granger causality testing, rolling-window stability analysis, and the newly proposed structural demand integration (SDI) index. By using this framework, three gaps in the state-of-the-art (SOTA) literature are addressed: the lack of structural (as opposed to predictive) Google Trends network studies in emerging Mediterranean contexts; the confounding of seasonal artifacts with structural co-movement in previous construction of networks; and the lack of evidence of dynamic demand leadership at a keyword level, analyzing the destination.

The three main empirical findings from the study are also extremely robust in relation to an extensive and substantial battery of validation tests. First, a vast concentration of tourist demand during the summer months is identified, specifically with Jale Beach being reported as having the highest summer ratio – 53.67 – of any Mediterranean tourism destination in the quantitative literature. The STL decomposition confirms that 10 from 15 of the keywords will exceed a seasonal F_s value of > 0.85, with only Llogara National Park (F_s = 0.874, summer ratio of 2.08) providing empirical evidence of a year-round anchor. Secondly, the raw correlation network, as initially constructed, has an extremely high density (0.835) and is statistically significantly different from a comparative random network (null model) at the 0.001 level. Thus, providing a strong confirmation of H1. The de-seasonalised correlation network identifies that there are only 6 nodes in the adjusted core of cross-category structural co-movement. As the third part of the analytical framework presented in this research indicates, Kanina Castle is structurally peripheral across every method used in this research; it has a DC score of 0 in the raw Pearson correlation network, no bootstrap confirmed edges, does not show up in community detection, is isolated within both split-sample time periods, and has no significant betweenness centrality in the permutation null model. In addition, the partial correlation analysis shows a statistically significant negative partial correlation of -0.417 between Kanina Castle and restaurants vlore which suggests that competition between these two components is structural (not merely passive), while Granger causality provides evidence that there is one statistically significant FDR corrected causal signal (from Gjipe Beach) being received by Kanina Castle, with no support for Kanina Castle being a leading cause of demand; thus confirming that Kanina Castle is a structural receiver only within the demand system.

As stated by the SDI Index, this study produces a replicable diagnostic instrument converting destination competitiveness theory into an empirically-grounded, keyword-level ranking, with the three highest-rated components for developing integration (St. Mary's Monastery 100; restaurant Vlora 91.6; and Sazan Island 72.5), indicating to DMO-based investment priority; thus, structural repositioning at Kanina Castle (SDI = 0; therefore, cannot produce marketing investment integration benefits) is required prior to producing integration benefits from marketing investment. The Granger leadership analysis identifies Gjipe Beach, Sazan Island, and the Bay of Grama as the three-monthly demand leaders, providing DMOs with a cost-free and real-time early-warning system for capacity planning.

In addition to the structural diagnostic aspect of the framework, it demonstrates direct predictive utility. Using a Network-VAR specification combining the previously defined Granger leaders (Gjipe Beach, Sazan Island, Bay of Grama) as exogenous regressors resulted in a statistically significant (Diebold-Mariano DM = -3.11; p = 0.0019) improvement over the ARIMA baseline to achieve a 66.3% reduction in forecast error for restaurant Vlora and consistent improvement across core destination targets. The non-linear information theory method, Transfer Entropy, validated the directional causality flows identified in this research (TE values approaching 0.40 bits). With these findings, the SDI Index and the entire network diagnostic framework are moved from descriptive to predictive utility, filling the previously existing predictive-validated gap of correlation network analyses applied to a tourism context.

From a methodological standpoint, this study was able to demonstrate that FDR correction on the 62 apparent Granger-significant relationships reduced them to 18 (a 71% reduction); null model comparison was necessary for determining validity of topological claims regarding the network; and de-seasonalization will assist in determining what percentage of the network edges can be attributed to shared seasonal cycles (91% in this context). The procedural contributions to methodology provided in this study create a replicable template for digital demand diagnostics in data-scarce emerging destinations within the Eastern Mediterranean, Adriatic, and Balkan coastal region, where DMOs will rely on the only available source of demand intelligence; thus, Google Trends serves as that source for local DMOs. The framework's reliance on freely available search data is further supported by its validation against official statistics: the destination search index co-moves strongly with national arrivals, including after seasonal adjustment, which indicates that the approach is anchored to real tourism activity and can be applied in emerging destinations where fine-grained official data are unavailable.

The future direction of research in marine tourism, outdoor recreation/activities and religious cultural heritage should focus on extending the keywords used to delineate which aspects of marine, outdoor and religion tourism and/or recreation, the geographic unit of analysis being more closely aligned with Google Trends data; the comparisons will need to be checked using the SDI rankings against both administrative arrival statistics provided by the Albanian National Institute of Statistics (INSTAT) and hotel occupancy data from the Vlora region to show that SDI rankings are valid or the validity of SDI rankings should continue to be established through national data sets. In addition, a comparison of the SDI framework with locations in Saranda and Durrës, and with selected Mediterranean destinations in both Greece and Croatia, should be made to determine the degree to which the SDI can be generalized across multiple destinations.

NOMENCLATURE

ADF	Augmented Dickey–Fuller test (stationarity)
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
BC	Betweenness Centrality
BCa	Bias-Corrected and Accelerated (bootstrap interval)
BH	Benjamini–Hochberg (FDR correction procedure)

CC	Connected Component
CI	Confidence Interval
CV	Coefficient of Variation
DBN	Dynamic Bayesian Network
DC	Degree Centrality
DMO	Destination Management Organisation
EEMD	Ensemble Empirical Mode Decomposition
ER	Erdős–Rényi (random graph model)
ERGM	Exponential Random Graph Model
FDR	False Discovery Rate
GT	Google Trends
H1, H2, H3	Research Hypotheses 1 through 5
INSTAT	Institute of Statistics of Albania
IQR	Interquartile Range
KPSS	Kwiatkowski–Phillips–Schmidt–Shin test (stationarity)
LSTM	Long Short-Term Memory neural network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MIDAS	Mixed-Data Sampling regression
n	Sample size
NL	Net (Granger) Leadership
ORCID	Open Researcher and Contributor ID
PCA	Principal Component Analysis
Q	Modularity score (community detection)
r	Pearson correlation coefficient
RMSE	Root Mean Square Error
SD	Standard Deviation
SDI	Structural Demand Integration Index
SNA	Social Network Analysis
SOTA	State Of The Art
STL	Seasonal-Trend decomposition using Loess
TE	Transfer Entropy
UNDP	United Nations Development Programme
VAR	Vector Autoregression
VECM	Vector Error Correction Model

AUTHOR CONTRIBUTIONS

Conceptualization, I.B. and E.M.; methodology, I.B. and E.M.; software, E.P.; validation, E.M. and S.M.; formal analysis, E.P.; investigation, E.P. and S.K.; resources, S.K.; data curation, E.P.; writing—original draft preparation, E.P.; writing—review and editing, I.B., E.M., S.K., and S.M.; visualization, E.P.; supervision, I.B. and E.M.; project administration, I.B.; funding acquisition, I.B. All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTERESTS

The authors declare no conflict of interest associated with this publication.

ACKNOWLEDGMENT

Our sincere gratitude to the National Agency of Scientific Research and Innovation (NASRI) in Albania for its generous support, which enabled us to carry out this study. The financial support received for our project "Research on Development of Strategic Tourism Investments", based on Decision No. 6, dated 10.06.2024, "On the approval of the financing of Research and Development Projects, for the Year 2024", is responsible for the study's significant success.

DATA AVAILABILITY STATEMENT

The Google Trends data used in this study are publicly available at https://trends.google.com. The complete Python analytical workflow is available as a reproducible Google Colab notebook in a private GitHub repository that will be public upon acceptance of the paper.

REFERENCES

1. Choi, H., Varian, H. Predicting the Present with Google Trends. *Economic Record*, **2012**, *88*, 2-9.
2. Wu, D.C., Zhong, S., Wu, J., Song, H. Tourism and Hospitality Forecasting with Big Data: A Systematic Review of the Literature. *J. Hosp. Tour. Res.*, **2025**, *49*, 615–634.
3. Önder, I., Gunter, U. Forecasting Tourism Demand with Google Trends for a Major European City Destination. *Tour. Anal.*, **2016**, *21*, 203–220.
4. Bokelmann, B., Lessmann, S. Spurious patterns in Google Trends data—An analysis of the effects on tourism demand forecasting in Germany. *Tour. Manag.*, **2019**, 75, 1–12.
5. Havranek, T., Zeynalov, A. Forecasting tourist arrivals: Google Trends meets mixed-frequency data. Tour. Econ. **2021**, *27*, 129–148.
6. UNDP. *Albania's Path to Climate Resilience: The National Adaptation Plan*. Available online: https://www.undp.org/sites/g/files/zskgke326/files/2026-01/web_tourism.pdf (access date 2 April 2026).
7. Burlea-Schiopoiu, A., Ozuni, F. The Potential of Albanian Tourism Sector. *Sustainability* **2021**, *13*, 3928.

8. Krabokoukis, T., Polyzos, S. Analyzing the Tourism Seasonality for the Mediterranean Countries. *J. Knowl. Econ.*, **2023**, 15, 8053–8076.
9. Butler, R. Seasonality in tourism: Issues and implications. *Tour. Rev.*, **1998**, 53, 18–24.
10. Niavis, S., Kallioras, D. The Efficiency of Tourism Sector in EU Mediterranean Coastal Regions: The Effects of Seasonality and Spatiality of Demand. *REGION*, **2021**, *8*, 135–152.
11. Cebrián, E., Domenech, J. Addressing Google Trends inconsistencies. *Technol. Forecast. Soc. Change*, **2024**, *202*, 123318.
12. Cebrián, E., Domenech, J. Is Google Trends a quality data source? *Appl. Econ. Lett.*, **2023**, *30*, 811–815.
13. Tsiotas, D., Krabokoukis, T., Polyzos, S. Detecting Tourism Typologies of Regional Destinations Based on Their Spatio-Temporal and Socioeconomic Performance: A Correlation-Based Complex Network Approach for the Case of Greece. *Tour. Hosp.*, **2021**, 2, 113–139.
14. Ritchie, J.R.B., Crouch, G. I. *The Competitive Destination: A Sustainable Tourism Perspective*; CABI: Wallingford, UK, **2003**.
15. Dann, G.M.S. Anomie, ego-enhancement and tourism. *Ann. Tour. Res.*, **1977**, *4*, 184–194.
16. Crompton, J.L. Motivations for pleasure vacation. *Ann. Tour. Res.*, **1979**, *6*, 408–424.
17. Bangwayo-Skeete, P. F., Skeete, R. W. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tour. Manag.*, **2015**, *46*, 454–464.
18. Thompson, H. Tourism Demand and Wages in a General Equilibrium Model of Production. *Tour. Econ.*, **2016**, *22*, 81–91.
19. Jackman, M., Naitram, S. Research Note: Nowcasting Tourist Arrivals in Barbados – Just Google it! *Tour. Econ.*, **2015**, *21*, 1309–1313.
20. Dinis, G., Breda, Z., Costa, C., Pacheco, O. Google Trends in tourism and hospitality research: a systematic literature review. *J. Hosp. Tour. Technol.*, **2019**, *10*, 747–763.
21. Dergiades, T., Mavragani, E., Pan, B. Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tour. Manag.*, **2018**, *66*, 108–120.
22. Höpken, W., Eberle, T., Fuchs, M., Lexhagen, M. Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: the case of Åre, Sweden. *Inf. Technol. Tour.*, **2019**, *21*, 45–62.
23. Buhalis, D. Marketing the competitive destination of the future. *Tour. Manag.*, **2000**, *21*, 97–116.
24. Seok, H., Barnett, G.A., Nam, Y. A social network analysis of international tourism flow. *Qual. Quant.* **2021**, *55*, 419–439.
25. Casanueva, C., Gallego, Á., García-Sánchez, M.-R. Social network analysis in tourism. *Curr. Issues Tour.* **2016**, *19*, 1190–1209.
26. Yang, Y. Understanding tourist attraction cooperation: An application of network analysis to the case of Shanghai, China. *J. Destin. Mark. Manag.*, **2018**, *8*, 396–411.
27. Tang, X., Fong, L.H.N., So, A.S.-I. Toward a framework for perceived yuanfen in the accommodation service encounter: a grounded theory study. *Int. J. Contemp. Hosp. Manag.*, **2024**, *36*, 155–181.
28. Harfst, J., Syrbe, R., Kern, C., Wirth, P., Sandriester, J., Pstrocka-Rak, M., et al. Cultural Tourism and Governance in Peripheral Regions. *Int. J. Tour. Res.*, **2024**, *26*, e2733.

29. Diebold, F.X., Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
30. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.*, **2000**, *85*, 461–464.
31. Cannas, R. An Overview of Tourism Seasonality: Key Concepts and Policies. *ALMT* **2012**, *3*, 40-58.
32. Ferrante, M., Lo Magno, G.L., De Cantis, S. Measuring tourism seasonality across European countries. *Tour. Manag.*, **2018**, *68*, 220–235.
33. Ruggieri, G., Platania, M. Islands' Tourism Seasonality: A Data Analysis of Mediterranean Islands' Tourism Comparing Seasonality Indicators (2008–2018). *Sustainability* **2024**, *16*, 3674.
34. Fernández-Morales, A., Cisneros-Martínez, J.D., McCabe, S. Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. *Tour. Manag.*, **2016**, *56*, 172–190.
35. Benur, A.M., Bramwell, B. Tourism product development and product diversification in destinations. *Tour. Manag.* **2015**, *50*, 213–224.
36. Sou, J.P., Siu, R.C.S. Does diversification drive tourism demand? A structural change perspective: Evidence from a casino tourism destination. *Int. J. Tour. Res.* **2023**, *25*, 293–304.
37. Dixit, S.K., Prayag, G. Gastronomic tourism experiences and experiential marketing. *Tour. Recreat. Res.*, **2022**, *47*, 217–220.
38. Gecaj, M., Mustaqe, E., Kadiu, E. Public Support Schemes and Multifunctional Rural Development: The Case of Agritourism in Albania. *Tour. Hosp.* **2026**, *7*, 204.
39. Dwyer, L., Kim, C. Destination Competitiveness: Determinants and Indicators. *Curr. Issues Tour.*, **2003**, *6*, 369–414.
40. Fodness, D., Murray, B. Tourist information search. *Ann. Tour. Res.*, **1997**, *24*, 503–523.
41. Gretzel, U., Sigala, M., Xiang, Z., Koo, C. Smart tourism: foundations and developments. *Electron. Mark.*, **2015**, *25*, 179–188.
42. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.*, **1978**, *1*, 215–239.
43. Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I. STL: A seasonal-trend decomposition procedure based on Loess. *J. Off. Stat.*, **1990**, 6, 3–73.
44. Richards, G. Cultural tourism: A review of recent research and trends. *J. Hosp. Tour. Manag.*, **2018**, *36*, 12–21.
45. Timothy, D.J. *Cultural Heritage and Tourism: An Introduction; Channel View Publications*: Bristol, UK, **2011**.