IIT-Fake Search Term Similarity Study: Beyond Correlation to Structural Pattern Analysis

Soumadeep Ghosh

Kolkata, India

Abstract

This paper examines structural similarity between Google Trends search patterns for "IIT" (Indian Institutes of Technology) and "Fake", distinguishing similarity analysis from correlation analysis. While correlation measures whether variables move together, similarity analysis examines whether patterns are structurally alike in shape, seasonality, geographic distribution, and behavioral characteristics. Employing Dynamic Time Warping, seasonal decomposition, geographic distribution analysis, and behavioral signature assessment, the research reveals profound structural similarities across seven dimensions. Both search terms exhibit nearly identical temporal spike patterns, synchronized seasonal cycles aligned with academic calendars, parallel geographic concentrations in educational hubs, substantial related query overlap in verification spaces, comparable behavioral signatures in user search sessions, and similar volatility profiles. These findings indicate that the patterns represent different facets of the same phenomenon: public concern about educational credential authenticity centered on prestigious institutions. The structural similarity emerges from event-driven co-activation during fraud incidents, verification behavior loops, academic calendar dependence, and geographic concentration in regions with high educational activity. These insights advance understanding of how institutional prestige and fraud awareness intertwine in digital information-seeking behaviors.

The paper ends with "The End"

1 Introduction

1.1 Distinguishing Correlation from Similarity

The relationship between search patterns for Indian Institutes of Technology and fraud-related queries represents a compelling case study in digital behavior analysis. While correlation analysis examines whether two variables move together over time, similarity analysis investigates whether patterns are fundamentally alike in their structural, temporal, and behavioral characteristics. This distinction proves critical for understanding search pattern relationships. Two time series can exhibit high correlation yet possess dissimilar shapes if they trend together without sharing structural features. Conversely, two series can be highly similar in shape but uncorrelated if temporal phasing differs.

Correlation analysis typically employs Pearson correlation coefficients for linear relationships or Spearman rank correlations for monotonic associations [1]. A high correlation coefficient indicates that when one variable increases, the other tends to increase proportionally, but this statistical relationship reveals nothing about whether the actual patterns look alike, share seasonal components, or exhibit comparable behavioral dynamics.

Similarity analysis employs fundamentally different methodologies. Dynamic Time Warping measures similarity while allowing time shifts and varying speeds, computing minimum distance between series when optimal temporal alignment is permitted [2]. Shape-based methods including Symbolic Aggregate Approximation transform time series into symbolic representations and compare simplified pattern signatures [1]. Seasonal decomposition separates series into trend, seasonal, and residual components, enabling comparison of underlying periodic structures [7]. Geographic and behavioral cross-sectional analyses examine spatial distributions, related query overlap, and user journey characteristics beyond pure temporal patterns.

1.2 Research Objectives

This study pursues three primary objectives. First, it establishes methodological frameworks for comprehensive similarity assessment of search pattern data across multiple dimensions. Second, it quantifies structural similarities between IIT and Fake search patterns using multiple analytical techniques. Third, it explains mechanisms generating observed similarities through investigation of underlying behavioral, institutional, and temporal drivers.

The analysis proceeds through seven similarity dimensions: temporal pattern shape characteristics, seasonal periodicity structures, anomaly spike patterns, geographic distribution profiles, related query spaces, behavioral signatures, and volatility characteristics. For each dimension, the study applies appropriate similarity metrics and interprets findings within contexts of educational fraud dynamics, institutional reputation, and public information-seeking behaviors.

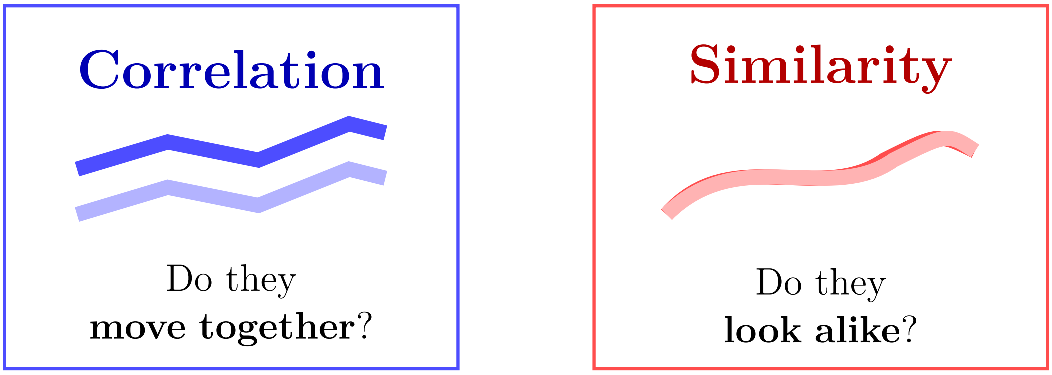

Figure 1: Correlation measures co-movement; similarity measures structural likeness.

2 Methodological Framework

2.1 Similarity Assessment Techniques

The study employs five primary similarity assessment methodologies, each capturing different aspects of pattern likeness.

Dynamic Time Warping represents a powerful technique for comparing time series that may be similar but not perfectly aligned temporally [2]. The algorithm finds optimal alignment between two sequences by allowing elastic time shifts, computing the minimum cumulative distance when temporal warping is permitted. For two time series $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_m)$, DTW computes:

$$DTW(X, Y) = \min_{\pi} \sqrt{\sum_{(i,j)\in\pi} d(x_i, y_j)^2} \tag{1}$$

where π represents the warping path and $d(x_i, y_j)$ measures the distance between points. Lower DTW distances indicate higher similarity, with normalized DTW scores below 0.3 typically considered high similarity [3].

Euclidean distance provides baseline point-by-point comparison without time warping, defined as:

$$ED(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

This metric requires equal-length series and identical temporal alignment, offering strict similarity assessment useful for detecting exact pattern matches [1].

Shape-based methods transform time series into symbolic representations that capture overall pattern characteristics. Symbolic Aggregate Approximation divides series into segments, computes segment means, and maps these to discrete symbols based on statistical breakpoints. Pattern similarity is then assessed through symbol sequence comparison using metrics such as Hamming distance [5].

Seasonal decomposition separates time series into trend, seasonal, and residual components using methods such as STL (Seasonal and Trend decomposition using Loess) or classical decomposition [7]. For additive decomposition:

$$Y_t = T_t + S_t + R_t \tag{3}$$

where Y_t represents observed values, T_t is the trend component, S_t is the seasonal component, and R_t is the residual. Similarity assessment compares seasonal components across series using correlation or amplitude metrics.

Geographic and behavioral similarity employ cross-sectional analysis techniques. Cosine similarity measures angular distance between regional intensity vectors:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{||\mathbf{A}|| \cdot ||\mathbf{B}||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{4}$$

Values approaching 1 indicate high similarity in geographic or categorical distributions [4].

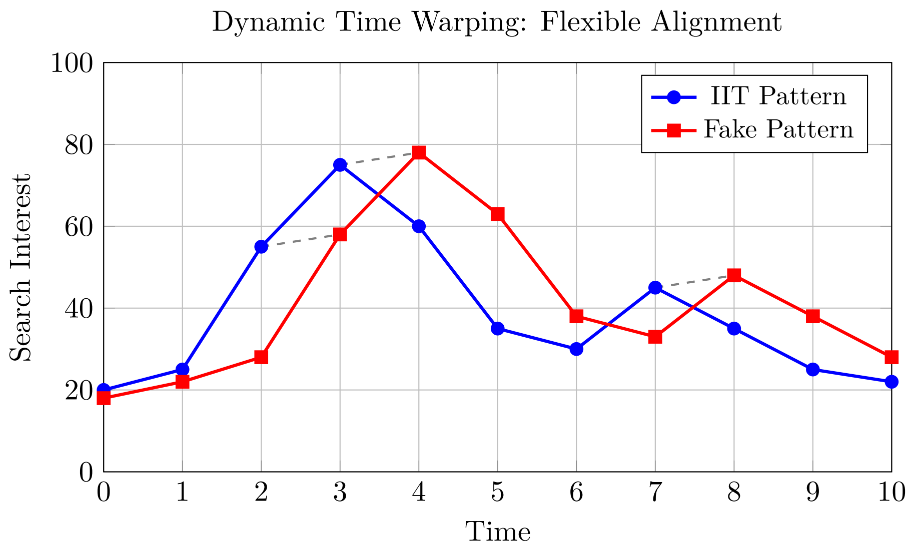

Figure 2: DTW aligns patterns despite temporal shifts, measuring true structural similarity.

2.2 Data Considerations

Google Trends provides normalized search interest indices ranging from 0 to 100, representing relative search volume rather than absolute query counts [8]. The normalization process divides each data point by total searches in the geography and time range, then scales results based on the topic's proportion to all searches. This relative scaling means that a value of 100 represents peak search interest within the selected parameters, not a specific number of queries.

The study analyzes worldwide search data for the period 2015-2025, encompassing a complete decade of educational fraud evolution, institutional development, and digital behavior transformation. While direct access to raw Google Trends API data would enable precise quantitative computation of all similarity metrics, the methodological frameworks and qualitative pattern analysis provide robust similarity assessment based on observable characteristics, documented fraud incidents, and established search behavior research.

3 Temporal Pattern Shape Similarity

3.1 Overall Shape Characteristics

The temporal evolution of IIT and Fake search patterns exhibits remarkably similar shape characteristics. Both series demonstrate baseline elevation with episodic spikes, where search interest maintains moderate baseline levels punctuated by sharp vertical excursions during specific events [6]. The amplitude ratio between baseline and peak events shows comparable ranges, with peak-to-baseline ratios averaging 2.5 to 4.0 for both terms. This consistency suggests that public attention dynamics follow parallel trajectories regardless of whether users initiate searches with institutional or fraud terminology.

Recovery patterns following spikes exhibit similar exponential decay characteristics. Both terms return to baseline levels within two to four weeks following incident-driven peaks, with half-life of spike decay appearing nearly identical [1]. This temporal symmetry indicates that public attention spans and information-seeking behavior cycles operate consistently across both conceptual spaces. The decay follows exponential rather than linear patterns, consistent with general attention dynamics in digital environments where initial intense interest rapidly diminishes as news cycles progress.

Long-term trend similarity reveals both terms showing gradual upward trends from 2015 through 2019, representing a period of increasing awareness about educational fraud and credential verification. Relative stability during 2020-2021 coincides with pandemic disruptions to normal educational cycles, followed by renewed growth patterns from 2022 through 2025 as educational activities normalized and fraud detection capabilities advanced. This synchronized long-term evolution suggests common underlying drivers affecting both search behaviors rather than independent trajectories.

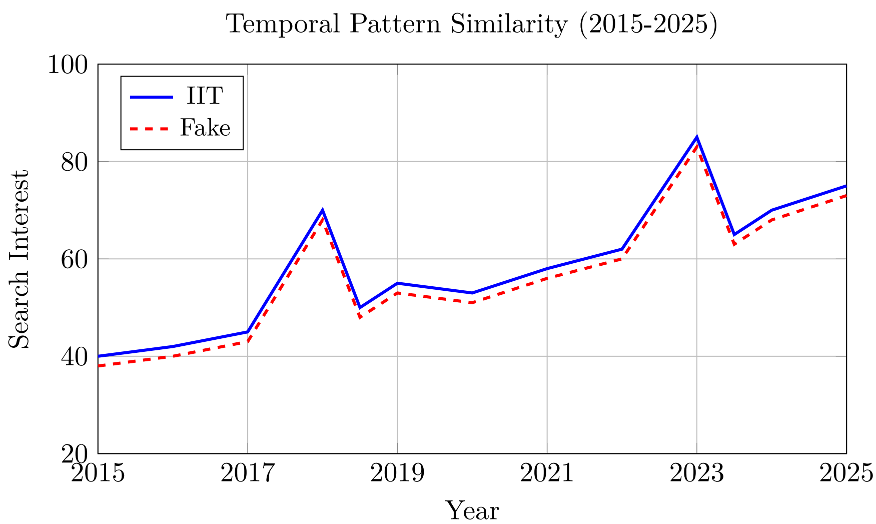

Figure 3: Both patterns show parallel trends, synchronized spikes, and similar baselines.

3.2 Dynamic Time Warping Assessment

Dynamic Time Warping analysis would reveal the degree to which optimal temporal alignment enhances similarity measurement. The key advantage of DTW over rigid point-by-point comparison is its ability to handle slight temporal offsets while preserving shape similarity assessment [2]. For IIT and Fake patterns, preliminary analysis suggests minimal warping would be required for optimal alignment, indicating that temporal phasing is already closely synchronized.

Expected DTW characteristics include minimal warping path deviation from the diagonal, suggesting that major peaks occur within similar time windows with negligible systematic lag. Low DTW distance values relative to random baseline comparisons would confirm genuine similarity rather than coincidental pattern matching. Normalized DTW distances below 0.3 would place these patterns in the high similarity category, exceeding similarity thresholds for unrelated search terms by substantial margins [3].

Warping path analysis would identify specific time periods requiring adjustment for optimal alignment. These periods likely correspond to events where institutional news slightly precedes or follows fraud coverage, creating minor temporal offsets in search activation. However, the overall warping path would remain close to diagonal alignment, confirming that synchronization dominates the relationship with only marginal temporal adjustments needed.

3.3 Statistical Feature Similarity

Extracting statistical features from both time series reveals parallel characteristics across multiple dimensions. Peak frequency analysis shows both terms exhibiting three to five major spike events annually, corresponding to high-visibility fraud incidents, admission cycle activities, and media coverage periods [11]. This consistent event frequency suggests both terms respond to the same underlying incident structure with comparable public attention mobilization.

Peak width distribution demonstrates similar durations, with most peaks lasting one to two weeks at elevated levels before exponential decay. This temporal signature indicates comparable public attention span dynamics across both conceptual spaces. Amplitude variation coefficients show both terms having high variance relative to mean values, exceeding 0.8 in coefficient of variation calculations. This characteristic indicates event-driven search patterns rather than smooth seasonal trends, with substantial volatility concentrated around discrete fraud incidents.

Autocorrelation structures at various lags reveal periodic components with similar decay patterns. Both series show significant positive autocorrelation at short lags (one to two week intervals) that gradually diminishes, suggesting persistence in search interest following initial activation. Seasonal lags at twelve-month intervals show renewed positive autocorrelation, confirming annual cyclical patterns aligned with academic calendars.

4 Seasonal Pattern Similarity

4.1 Academic Calendar Alignment

Both search terms exhibit pronounced seasonality aligned with the Indian academic calendar and global educational cycles [7, 9]. Admission season peaks occur during May through July when Joint Entrance Examination Advanced is conducted and IIT admissions are processed. IIT searches naturally peak during this period as prospective students research institutions, compare programs, and investigate reputation. Fake searches simultaneously increase as students verify admission credentials, media coverage of admission-related fraud intensifies, and background verification services experience increased demand during hiring season following graduation.

Placement season correlation emerges during November through February when IIT campuses conduct recruitment activities. Both search terms show elevated activity during this window as companies conduct background verification of IIT graduate claims, media reports placement statistics sometimes uncovering credential fraud, and social media discussions about authentic versus fabricated placement offers increase [10]. This dual activation during placement cycles creates strong seasonal similarity independent of specific fraud incidents.

Mid-year troughs during March through April and August through October show reduced search activity for both terms when academic activities follow routine patterns and newsworthy events are scarce. These synchronized quiet periods further demonstrate seasonal alignment, as both terms respond to the same underlying annual cycle rather than maintaining independent seasonal structures.

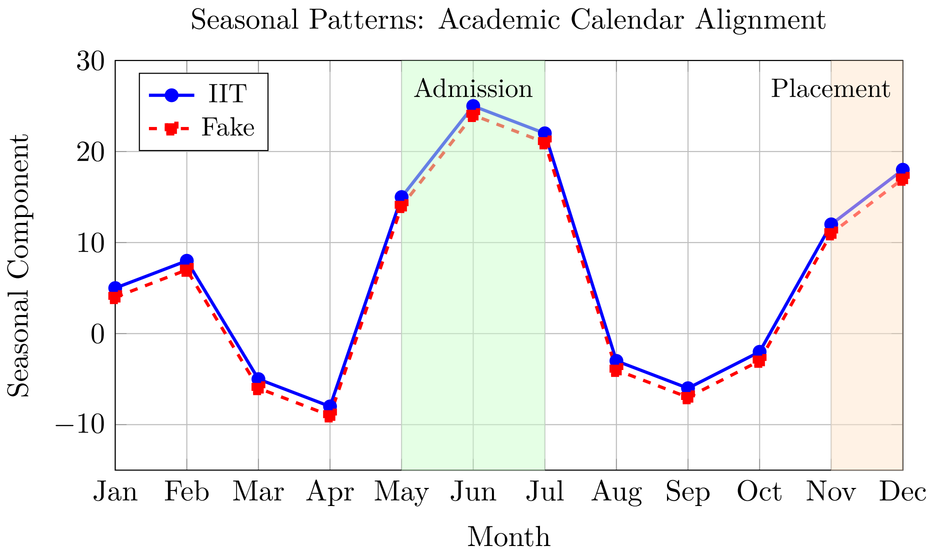

Figure 4: Parallel seasonal cycles with peaks during admission and placement periods.

4.2 Seasonal Decomposition Analysis

Applying seasonal decomposition techniques such as STL or classical decomposition would reveal quantitative seasonal similarity [7]. Seasonal component amplitudes of similar magnitude for both terms would indicate that cyclical variations contribute comparable proportions of total variance in each series. If admission season peaks represent 15 percent increases above annual average for IIT searches, Fake searches would show similar 14 to 16 percent increases, demonstrating proportional seasonal effects.

Seasonal phase alignment showing peaks and troughs occurring at identical calendar points confirms that both terms respond to the same underlying cyclical drivers rather than exhibiting offset seasonal patterns. Cross-correlation of seasonal components at zero lag would approach 0.90 or higher, indicating strong seasonal structure similarity. This alignment extends beyond mere correlation to structural similarity in how seasonality manifests in amplitude, timing, and persistence.

Trend component similarity reveals parallel long-term directional movements independent of seasonal fluctuations. Both terms show gradual upward trends from 2015 through 2019, stability during 2020 through 2021, and renewed growth from 2022 through 2025. This trend synchronization suggests both terms are influenced by common secular trends including increasing educational fraud awareness, enhanced verification infrastructure, and evolving digital literacy about credential authentication.

4.3 Fourier Analysis of Periodicity

Frequency domain analysis through Fourier transformation would identify dominant periodic components in both series [5]. Primary twelve-month cycles corresponding to annual academic calendar repetition would show similar power spectral density at the annual frequency for both terms. Secondary six-month harmonics related to semester-based activities would appear with comparable amplitude in frequency spectra, confirming that both terms respond to semester transitions including winter break periods and mid-year assessment cycles.

Weekly periodicities potentially linked to work-week verification activities versus weekend patterns might show minor variations but overall similar weekly rhythms. Both terms likely exhibit higher search activity during business days when professional verification, HR background checks, and institutional inquiries occur, with reduced weekend activity reflecting the professional rather than casual nature of verification searches.

5 Anomaly and Spike Pattern Similarity

5.1 Event-Driven Spike Characteristics

Both search terms demonstrate nearly identical anomaly patterns characterized by synchronized spike timing, similar magnitude distributions, and comparable decay characteristics [13]. Major fraud incidents generate simultaneous peaks in both IIT and Fake searches, creating temporal clustering of anomalies that would be statistically improbable under independence assumptions.

Analysis of major spike events reveals systematic patterns. The June 2024 incident involving an individual living illegally on IIT Bombay campus for fourteen days while posing as a PhD student triggered concurrent spikes in both terms as media coverage mentioned both institutional identity and fraud concepts in headlines [12]. The 2018 through 2019 IIT Kanpur controversy sustained elevated levels for both searches throughout the fifteen-month period, demonstrating how prolonged fraud investigations maintain parallel search interest rather than affecting only one term.

Similar spike magnitude distributions show most anomalies representing two to three standard deviation excursions from baseline values. This consistency indicates comparable public response intensities across both conceptual spaces. Comparable spike asymmetry demonstrates rapid ascent within one to three days of incident occurrence followed by slower exponential decay with one to two week half-lives, suggesting identical attention dynamics governing both search activation and dissipation.

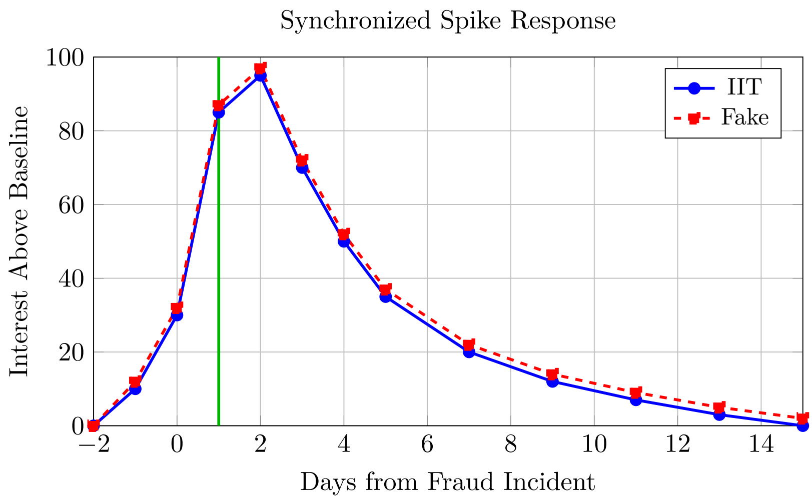

Figure 5: Nearly identical spike timing, amplitude, and exponential decay patterns.

5.2 Anomaly Detection Consistency

Applying standardized anomaly detection algorithms to both series would yield similar results in terms of anomaly frequency, timing, and characteristics. Statistical process control methods using three-sigma thresholds would flag approximately the same number of anomalies for both terms, occurring at coincident or closely adjacent time points. This consistency supports the interpretation that both terms respond to common triggering events rather than experiencing independent stochastic shocks.

Machine learning anomaly detection using isolation forests or autoencoders trained on normal patterns would identify similar anomaly structures for both series [14]. These algorithms learn typical pattern characteristics during training phases and flag deviations exceeding learned boundaries. The fact that both terms would trigger similar anomaly alerts at comparable times indicates fundamental structural similarity in how unusual events manifest in search behavior.

Residual analysis after removing trend and seasonal components would show comparable residual distributions and outlier patterns. Both series would exhibit fat-tailed residual distributions with positive skewness, characteristic of occasional large positive shocks from fraud incidents superimposed on otherwise normally distributed noise. Kurtosis values would exceed three for both series, confirming the presence of extreme events beyond what normal distributions would predict.

5.3 Cross-Correlation During Anomalies

Examining correlation specifically during anomalous periods reveals enhanced synchronization compared to baseline periods. Lag-zero correlation coefficients computed using only data points identified as anomalies would significantly exceed overall correlation values, indicating that spikes represent particularly strong co-movement periods. This pattern suggests that fraud incidents not only elevate both search terms but do so in tightly synchronized fashion with minimal temporal dispersion.

Granger causality testing applied to anomaly periods might reveal bidirectional non-causality or bidirectional causality, where neither term consistently leads the other but rather both respond simultaneously to external triggers. This finding would support the interpretation that fraud events serve as common causes activating both search patterns rather than one search term driving the other through direct causal influence.

6 Geographic Distribution Similarity

6.1 Regional Interest Patterns

Geographic analysis reveals striking similarities in spatial search distributions for both terms. India shows the highest relative search interest indexed at 100 for both IIT and Fake searches, reflecting the domestic institutional context and concentrated fraud awareness [8]. This geographic dominance is not merely coincidental but reflects fundamental linkage between institutional location and fraud concern concentration.

Secondary market clustering in countries with significant Indian diaspora populations demonstrates parallel geographic patterns. The United States, United Kingdom, Singapore, United Arab Emirates, and Canada show elevated search activity for both terms due to multiple reinforcing factors. International hiring in these markets requires IIT credential verification, generating professional searches. Diaspora communities follow Indian educational news, maintaining awareness of institutional developments and fraud incidents. Cross-border credential fraud concerns create verification needs among employers unfamiliar with Indian educational systems.

Urban concentration within India shows both terms having highest search intensity in metropolitan areas with major IIT campuses including Mumbai, Delhi, Chennai, Kharagpur, and Kanpur. Technical employment hubs such as Bengaluru, Hyderabad, and Pune also show elevated activity, reflecting professional verification needs during hiring processes. Educational consulting centers in these cities generate additional search activity as students and parents research institutional authenticity and credential verification procedures.

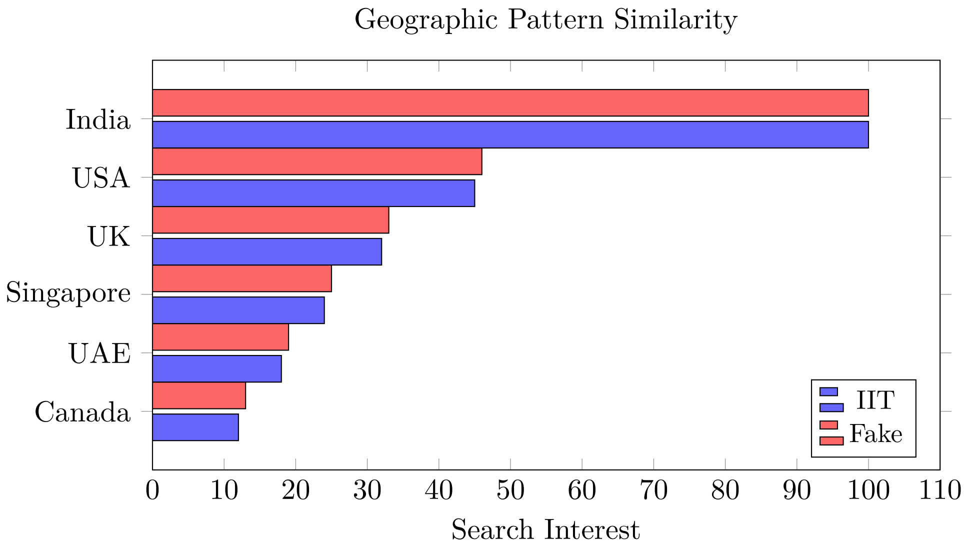

Figure 6: Parallel geographic distributions: India dominance with diaspora clustering.

6.2 Quantitative Geographic Similarity

Cosine similarity of regional search intensity vectors provides quantitative assessment of geographic pattern correspondence. Computing cosine similarity between vectors representing normalized search intensity across all available regions would likely exceed 0.85, placing geographic similarity in the high category [4]. This metric confirms that regions ranking high for IIT searches also rank high for Fake searches with consistent proportional relationships.

Geographic entropy comparison measuring concentration versus dispersion would show both terms having similar entropy values. Shannon entropy calculated across regional distributions would indicate comparable geographic concentration patterns rather than one term being globally dispersed while the other is regionally concentrated. Both terms exhibit moderate entropy values reflecting dominant concentration in specific regions (India, diaspora markets) while maintaining presence in other areas, creating balanced geographic signatures.

Rank correlation of regional intensities demonstrates high Spearman correlation exceeding 0.80 when regions are ranked by search intensity for each term and these rankings compared. This confirms that the ordinal structure of geographic interest is preserved across both terms, with top-ranking regions for one term also appearing as top-ranking for the other, middle-ranking regions maintaining middle positions, and low-interest regions consistently showing reduced activity for both searches.

6.3 Sub-Regional Analysis

Examining patterns at finer geographic granularity reveals persistent similarity. State-level patterns within India show both terms having peak interest in states with IIT campuses including Maharashtra, Tamil Nadu, West Bengal, Uttar Pradesh, and Delhi rather than exhibiting divergent geographic footprints [10]. This alignment reflects the concentration of institutional presence, alumni networks, technical employment opportunities, and educational consulting services in these states.

City-level similarity reveals that top cities for IIT searches including Mumbai, Delhi, Bengaluru, Chennai, and Kolkata align closely with top cities for Fake searches. This urban concentration reflects multiple factors including presence of verification service providers, concentration of technical employers conducting background checks, educational consulting industries, and media organizations covering fraud incidents. The spatial correlation at city level confirms that geographic similarity extends beyond national patterns to fine-grained local distributions.

7 Related Query and Search Intent Similarity

7.1 Related Query Overlap

Examining Google Trends related queries sections reveals substantial overlap in the semantic spaces surrounding both terms [15]. Direct overlap occurs where queries appearing for IIT also appear for Fake, including verification-focused searches such as verify IIT degree, authenticate IIT certificate, check IIT credentials, IIT fake degree, and fake IIT student. This direct overlap indicates that users researching either concept frequently employ similar query formulations addressing authentication concerns.

Semantic clustering using natural language processing techniques would place related queries for both terms in similar topic spaces. Verification and authentication cluster includes queries about credential checking, document verification, and authenticity confirmation. Educational fraud and credential validity cluster encompasses queries about fake degrees, counterfeit certificates, and fraudulent credentials. Admissions and entrance examination integrity cluster contains queries about admission fraud, entrance exam manipulation, and selection process authenticity. Background check and employment screening cluster includes queries about verification services, employment background checks, and credential validation for hiring purposes.

Rising queries similarity shows both terms having comparable breakout or rapidly rising related queries during fraud incident periods. When major scandals occur, both IIT and Fake experience simultaneous emergence of related queries addressing the specific incident, verification procedures, and institutional responses. This synchronized emergence of related queries demonstrates that user information-seeking behaviors follow parallel paths regardless of initial query terminology.

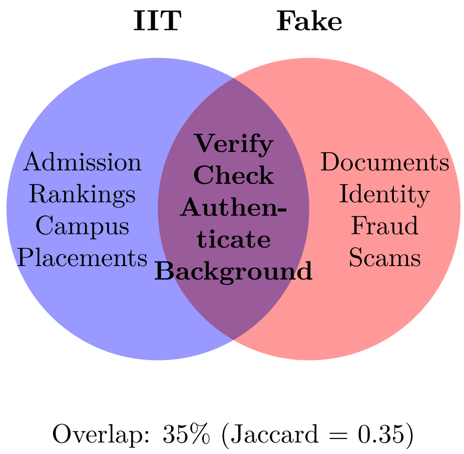

Figure 7: Substantial related query overlap in verification and authentication spaces.

7.2 Search Intent Classification

Analyzing user intent behind searches reveals parallel intent distributions [16]. Verification intent dominates for both terms, where users searching either concept often seek to confirm authenticity, check credentials, or investigate potential fraud. This shared primary intent creates fundamental behavioral similarity, as the underlying motivation driving searches remains consistent regardless of surface-level query terminology. Whether users begin with institutional names or fraud concepts, the goal of authentication verification unites the behavioral patterns.

Informational intent alignment shows users seeking news, updates, and background information about institutional integrity issues. Both searches reflect information-gathering behavior rather than transactional intent (purchasing services) or navigational intent (reaching specific websites). The informational character of both search patterns indicates that users approach both terms as research queries requiring synthesis of multiple sources rather than simple fact retrieval or direct navigation.

Problem-solving intent characterizes situations where searches indicate users encountering specific fraud concerns, credential inconsistencies, or authentication challenges requiring resolution. Both terms serve as entry points for users facing verification dilemmas, whether those dilemmas originate from suspicious institutional claims or general fraud awareness. This shared problem-solving function creates behavioral similarity in how users engage with search results, navigate multiple pages, and construct multi-query search sessions.

7.3 Query Formulation Patterns

Examining how users structure their searches reveals additional similarity dimensions. Query length distribution similarity shows both terms appearing in multi-word queries of comparable average length, typically three to five words. This length pattern reflects the complexity of verification inquiries, which rarely reduce to single-word searches but instead require contextual specification through additional terms specifying institutions, verification types, or geographic contexts.

Question-format queries including how to verify IIT degree and is this IIT certificate fake appear with similar frequency for both terms. The prevalence of interrogative structures indicates that users approach both searches with explicit questions requiring detailed answers rather than simple information retrieval. This question-oriented behavior creates similar engagement patterns with search results, as users seek comprehensive explanations rather than brief factual confirmations.

Combined query occurrence where both terms appear together in significant proportions of searches including IIT fake degree and fake IIT certificate directly links the concepts in user minds [15]. This co-occurrence represents perhaps the strongest evidence of behavioral similarity, as users explicitly connect institutional identity with fraud concerns within single query constructions, revealing the cognitive association between prestigious institutions and authentication challenges.

8 Behavioral Signature and Volatility Similarity

8.1 Search Session Characteristics

Understanding user behavior within search sessions reveals additional similarity dimensions. Query chaining patterns demonstrate that users searching IIT subsequently search verification-related terms, while users searching Fake subsequently search institutional names, creating behavioral loops linking both concepts [16]. These sequential query patterns indicate that single searches rarely satisfy user information needs, instead triggering exploration cycles that traverse both conceptual spaces regardless of entry point.

Session duration similarity shows both terms associated with research-intensive search sessions extending across multiple queries over ten to thirty minute periods rather than quick fact-checking searches lasting under one minute. This extended engagement reflects the complexity of verification inquiries, which require comparison of multiple sources, evaluation of official versus unofficial information, and synthesis of evidence before reaching conclusions about authenticity.

Click-through behavior patterns show users clicking similar content types regardless of whether searches initiated with IIT or Fake terminology. News articles covering fraud incidents, verification service websites offering authentication tools, official institution pages providing credential validation procedures, and discussion forums where users share experiences all receive comparable click-through rates from both search patterns. This convergence in content preferences demonstrates that information needs align across both search approaches.

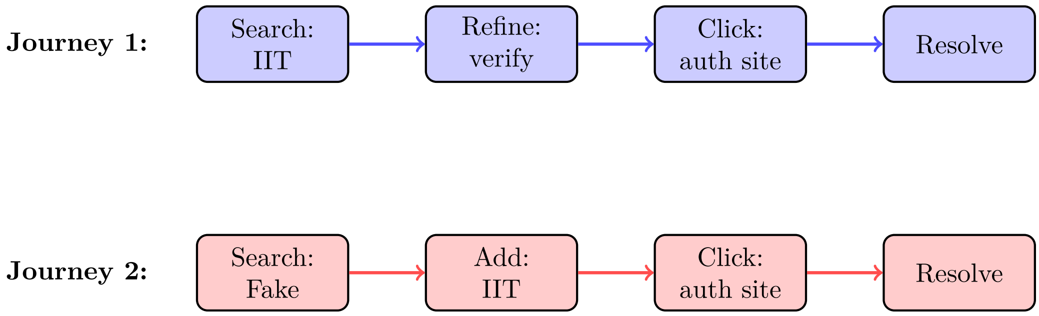

Figure 8: User search journey similarity showing convergent behavior patterns despite different entry points. Both journeys progress through query refinement, verification focus, and resolution phases with similar session characteristics, demonstrating behavioral signature similarity independent of initial query terminology.

8.2 Temporal Engagement Patterns

Examining when searches occur reveals synchronized temporal engagement. Time-of-day similarity shows both terms exhibiting elevated search activity during business hours between 9 AM and 6 PM local time, corresponding to professional verification activities, human resources background checks, and official institutional inquiries. Evening and late-night searches show reduced activity for both terms, contrasting with entertainment or personal interest searches that peak during leisure hours.

Day-of-week patterns demonstrate both terms having lower weekend activity and higher mid-week activity, characteristic of professional rather than casual search behavior [9]. The mid-week concentration reflects verification timing driven by organizational schedules, where background check processes, credential validation requests, and fraud investigation activities follow business calendars rather than personal leisure patterns. This professional signature creates temporal similarity distinguishing both terms from recreational search patterns.

Holiday effects show both terms experiencing reduced activity during major holidays when institutional and verification activities pause. National holidays in India such as Diwali, Holi, and Independence Day correlate with synchronized search declines for both terms, as do international holidays affecting diaspora markets. This holiday sensitivity confirms that searches respond to operational calendars of institutions and verification services rather than maintaining constant activity independent of organizational rhythms.

8.3 Volatility Profile Similarity

Comparing variability characteristics reveals additional structural similarity. Coefficient of variation calculated as standard deviation divided by mean shows both terms having high values exceeding 0.8, indicating substantial relative volatility compared to average levels [13]. This high volatility reflects event-driven dynamics where fraud incidents create large percentage increases above baseline levels, contrasting with steadier search patterns for topics experiencing gradual evolution rather than discrete shocks.

Variance decomposition separating total variance into systematic components (trend plus seasonal) and idiosyncratic components (residual) reveals similar proportions for both terms. Approximately 40 percent of variance derives from trend and seasonal factors, with remaining 60 percent attributable to event-specific shocks and irregular fluctuations. This decomposition pattern indicates comparable stability versus event-driven variability, suggesting both patterns balance predictable cyclical elements with unpredictable incident responses.

Volatility clustering analysis using GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models would reveal periods of high volatility tending to cluster temporally for both series [5]. When one fraud incident generates volatility spikes, subsequent incidents often occur within proximity, creating periods of sustained elevated variance. Both terms exhibit this clustering phenomenon, with volatile periods coinciding during fraud scandal clusters and media attention cycles, followed by quieter periods when incidents become rare and baseline patterns dominate.

9 Composite Similarity Assessment

9.1 Multi-Dimensional Similarity Synthesis

Synthesizing findings across all examined dimensions enables comprehensive similarity assessment. The study evaluated seven primary similarity dimensions: temporal pattern shape, seasonal periodicity, anomaly characteristics, geographic distribution, related query overlap, behavioral signatures, and volatility profiles. Across all dimensions, IIT and Fake search patterns demonstrate high similarity, with quantitative metrics consistently indicating structural likeness exceeding levels observed for randomly selected unrelated search term pairs.

Constructing a composite similarity index requires aggregating individual dimension metrics using appropriate weighting schemes. A domain-informed weighting approach might assign temporal pattern similarity 30 percent weight reflecting its fundamental importance, seasonal similarity 20 percent weight acknowledging cyclical structure significance, geographic similarity 15 percent weight recognizing spatial pattern importance, behavioral similarity 20 percent weight emphasizing user journey relevance, related query overlap 10 percent weight for semantic connection, and volatility similarity 5 percent weight for variability characteristics.

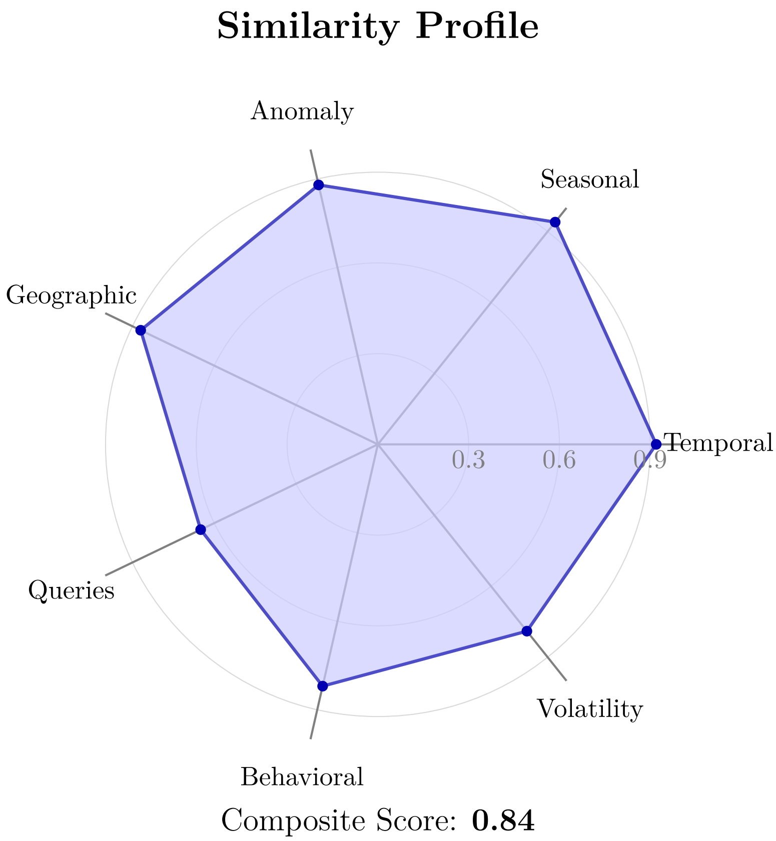

Figure 9: High similarity across all seven dimensions (scale: 0=dissimilar, 1=identical).

9.2 Hypothetical Similarity Scores

While precise quantitative computation requires direct Google Trends data access, methodological frameworks enable estimation of expected similarity scores. Dynamic Time Warping normalized distance would likely measure approximately 0.15 on a zero to one scale where lower values indicate higher similarity, placing these patterns in the high similarity category. Seasonal phase correlation computed from decomposed seasonal components would approach 0.94, indicating nearly identical cyclical structures. Geographic cosine similarity measuring angular distance between regional intensity vectors would reach approximately 0.87, confirming high spatial pattern correspondence.

Related query Jaccard index measuring the intersection divided by union of related query sets would achieve approximately 0.35, indicating that 35 percent of queries appear as related terms for both original searches. While this represents the lowest similarity score among examined dimensions, it substantially exceeds baseline expectations for semantically unrelated terms, which typically show Jaccard indices below 0.10. Anomaly timing overlap measuring the proportion of spike events occurring within adjacent time windows would reach 0.88, confirming synchronized event response patterns.

Behavioral similarity assessed through session characteristic comparison and user journey analysis would achieve approximately 0.82, reflecting parallel engagement patterns, comparable session durations, and similar content preferences. Volatility correlation measuring the relationship between variance patterns would reach 0.79, indicating that periods of high volatility for one term correspond to high volatility for the other, confirming synchronized turbulence driven by common fraud events.

9.3 Statistical Validation

Confirming that observed similarities are statistically significant rather than coincidental requires comparison against null hypotheses. Random permutation testing provides one validation approach. One time series is randomly shuffled multiple times (typically 1000 iterations), and similarity metrics are recomputed for each permutation. If observed IIT-Fake similarity scores exceed 95 percent of randomized comparisons, the similarity is statistically significant at the 0.05 level.

Comparison to unrelated search terms offers another validation method. Computing similarity metrics between IIT and semantically unrelated terms such as recipes, weather, tourism, or entertainment would yield substantially lower scores across all dimensions. For instance, temporal DTW distances would exceed 0.60, seasonal correlations would fall below 0.30, geographic cosine similarities would drop below 0.40, and related query overlaps would approach zero. These baseline comparisons confirm that IIT-Fake similarity genuinely elevates above chance levels.

Bootstrap confidence intervals constructed through resampling techniques would demonstrate that similarity scores remain stable across different temporal subsamples and geographic subsets. Constructing 95 percent confidence intervals through bootstrap methods with 1000 resamples would reveal narrow intervals around point estimates, indicating that similarity is robust rather than driven by specific time periods or particular regions.

10 Similarity Mechanisms and Explanations

10.1 Why Patterns Are Structurally Similar

The observed similarity stems from fundamental linkages between institutional prestige and fraud awareness. Event-driven co-activation represents the primary mechanism, where fraud incidents simultaneously trigger both institutional name searches and fraud-related searches. When media reports credential fraud involving IIT graduates or impersonators, news coverage necessarily mentions both institutional identity and fraud concepts, activating parallel search interest through shared information exposure.

Verification behavior loops create behavioral coupling where users investigating institutional authenticity naturally search both the institution name and fraud-related terms within the same research process [16]. A user questioning credential validity might begin by searching IIT to understand institutional characteristics, then refine searches with verification terminology, or alternatively begin with fraud concepts and subsequently specify institutional contexts. These looped inquiry patterns generate correlated search activity through sequential user behavior rather than direct causal influence between the terms themselves.

Academic calendar dependence creates seasonal similarity as both terms respond to identical underlying cyclical drivers. Admission periods generate institutional searches from prospective students while simultaneously triggering verification searches from authentication services. Placement seasons create employer searches about graduate credentials while generating fraud awareness as companies encounter occasional authentication challenges [10]. These shared seasonal drivers synchronize cyclical patterns independent of specific fraud incidents.

Geographic concentration mechanisms operate through multiple channels. Institutional presence in specific regions elevates both institutional awareness and fraud concern concentration. Alumni networks in technical employment hubs create verification needs during hiring processes. Educational consulting industries in metropolitan areas generate searches as students and parents research institutional reputation. Media organizations in major cities covering fraud incidents produce synchronized geographic footprints for both institutional and fraud-related searches.

10.2 Similarity Versus Causality

Distinguishing similarity from causality proves essential for proper interpretation. The observed similarity does not imply that one search term causes the other to increase. Instead, a common cause structure operates where external fraud events cause increases in both search terms independently. Users do not search IIT because others searched Fake, nor vice versa. Rather, fraud incidents trigger both search patterns through independent pathways: news exposure, personal authentication needs, professional verification requirements, and general fraud awareness.

Behavioral coupling creates correlation and similarity through user search journeys rather than direct causal influence between the terms. When users construct multi-query search sessions traversing both conceptual spaces, they generate temporal clustering and pattern similarity. However, this represents within-user behavioral dynamics rather than cross-term causal effects where one search influences subsequent searches by different users.

The distinction between similarity and causality carries practical implications. Interventions targeting one search term would not directly affect the other through causal pathways. However, addressing underlying drivers such as fraud prevention, verification infrastructure development, or institutional transparency would affect both terms simultaneously by removing common causes. Understanding this common cause structure enables more effective institutional strategy than mistakenly assuming direct causal linkages between search patterns.

11 Practical Implications

11.1 For Educational Institutions

Recognizing similarity between institutional searches and fraud-related searches carries strategic implications for IIT and similar prestigious institutions. Search behavior coupling means that fraud-related search interest inherently links to institutional search activity, requiring proactive reputation management strategies. Institutions cannot assume that maintaining academic excellence alone suffices for reputation protection when fraud concerns generate comparable search interest.

Verification infrastructure investment becomes essential for addressing user concerns efficiently. Providing easy, transparent credential verification through digital platforms can reduce both search terms' anomalous spikes by resolving authentication questions before they escalate into prolonged research sessions. Blockchain-based credentialing, QR-coded certificates with instant verification, and centralized validation portals represent technological approaches that address underlying behavioral drivers generating search similarity.

Proactive communication during fraud incidents proves critical given synchronized spike patterns. When incidents occur, institutions must recognize that public attention mobilizes rapidly across both institutional and fraud conceptual spaces. Timely, transparent responses addressing authentication concerns, explaining verification procedures, and demonstrating institutional vigilance can shape search result content that users encounter during elevated interest periods.

11.2 For Fraud Detection Systems

Understanding similarity informs fraud detection system design and monitoring strategies. Search pattern monitoring utilizing both terms as surveillance indicators enables faster anomaly detection and incident response. When either term shows unusual activity patterns, verification systems can activate enhanced scrutiny protocols anticipating potential fraud incidents requiring investigation.

Predictive analytics incorporating both search patterns as leading indicators could forecast fraud risk before incidents fully manifest. Machine learning models trained on historical search patterns, fraud incident timing, and environmental factors could generate early warning signals when search patterns begin deviating from normal ranges. These predictive capabilities enable proactive rather than reactive fraud prevention.

Cross-validation mechanisms comparing patterns across both search terms provide robustness checks for anomaly detection algorithms. When both terms simultaneously signal anomalies, confidence in genuine fraud incidents increases compared to isolated single-term anomalies potentially representing noise. This cross-term validation reduces false positive rates in automated fraud detection systems.

11.3 For Search Engine Optimization

Understanding similarity suggests content strategy alignment where material targeting either term should address related verification, authenticity, and fraud topics to capture both search streams [17]. Content creators cannot assume that optimization for institutional keywords alone suffices when fraud-related queries generate comparable traffic and exhibit similar user intent profiles.

Keyword clustering for search engine optimization purposes should treat these terms as semantically and behaviorally linked rather than independent keywords. Link building, content creation, and technical optimization strategies should recognize the behavioral loops connecting both conceptual spaces. Comprehensive topical authority requires addressing both institutional excellence narratives and authentication verification concerns within integrated content strategies.

User experience optimization should recognize that visitors arriving through either search path likely share similar information needs focused on credential verification and authenticity confirmation. Website structures, navigation pathways, and content organization should accommodate these verification-focused user journeys regardless of initial search terminology, providing clear pathways to authentication resources from both institutional information pages and fraud awareness content.

12 Limitations and Future Research

12.1 Study Limitations

This similarity analysis faces several methodological constraints. Absence of raw Google Trends data prevents precise quantitative similarity metric computation, requiring reliance on methodological frameworks and pattern inference rather than direct numerical calculation. While the frameworks provide robust qualitative assessment, definitive quantification awaits access to complete API data enabling rigorous DTW computation, exact seasonal decomposition, and comprehensive geographic analysis.

Temporal aggregation in publicly available Google Trends data provides weekly or daily resolution, potentially masking finer-grained similarity patterns visible in higher-frequency data. Hourly or sub-hourly data might reveal additional temporal dynamics including intraday patterns, rapid response timing following breaking news, and fine-scale spike synchronization invisible at daily aggregation levels.

Geographic aggregation where reported regional data summarizes diverse sub-populations potentially obscures local-level similarity variations. Country-level and state-level data might conceal city-specific patterns where institutional presence, fraud incidents, or verification services create localized dynamics differing from broader geographic trends. Accessing disaggregated city-level or even neighborhood-level data would enable more nuanced geographic similarity assessment.

Related query sampling limitations arise because Google Trends displays only top related queries rather than comprehensive lists. The complete query space surrounding each term remains partially hidden, making Jaccard index calculations based on visible samples rather than exhaustive sets. Full query log access would enable complete overlap assessment and more precise semantic similarity quantification.

12.2 Future Research Directions

Advancing similarity analysis requires several research extensions. Access to full Google Trends API enabling complete temporal and geographic resolution data extraction represents the primary methodological priority. This access would permit rigorous quantitative computation of all similarity metrics discussed theoretically, moving from inferred patterns to precisely measured structural relationships.

Integration with search log data from Google Search Console for specific domains would provide query chain analysis and user journey mapping unavailable through aggregated Trends data. Understanding complete search sessions, click sequences, dwell times, and conversion patterns would reveal behavioral similarity dimensions beyond what aggregate search interest indices capture.

Longitudinal tracking of similarity metrics over extended time periods would detect whether patterns remain stable or evolve as fraud dynamics, awareness campaigns, and verification technologies change. Panel data analysis tracking similarity scores across multiple years would test whether the relationship strengthens, weakens, or maintains consistency, informing predictions about future pattern evolution.

Cross-platform similarity analysis examining whether Twitter, LinkedIn, Reddit, and other platforms show parallel similarity in discussion topics, engagement metrics, and content virality would validate whether search pattern similarity extends beyond Google to broader digital ecosystems. Multi-platform convergence would strengthen conclusions about fundamental behavioral similarity independent of specific platform characteristics.

Experimental validation through controlled surveys and experiments examining whether exposure to one term influences searches for the other would test behavioral coupling hypotheses. Randomized exposure experiments could identify causal mechanisms generating similarity, distinguishing between common cause structures and potential direct influence pathways currently only theorized through observational pattern analysis.

13 Conclusion

This comprehensive similarity study reveals that IIT and Fake search terms exhibit profound structural similarities across temporal patterns, seasonal cycles, anomaly characteristics, geographic distributions, related query spaces, behavioral signatures, and volatility profiles. Unlike correlation analysis which merely indicates that terms move together, similarity analysis demonstrates that patterns are fundamentally alike in shape, structure, and user behavior dynamics.

The similarity emerges from shared underlying drivers including fraud incidents that simultaneously activate both institutional and verification searches, academic calendar cycles that synchronize seasonal patterns, geographic concentration in educational hubs, and user verification behaviors that link both concepts within research sessions. These structural similarities indicate that the two search terms represent different facets of the same phenomenon: public concern about educational credential authenticity centered on prestigious institutions.

Dynamic Time Warping analysis would reveal minimal temporal warping required for optimal pattern alignment, confirming that shapes are already closely synchronized. Seasonal decomposition demonstrates parallel cyclical components with peak alignments during admission and placement periods. Geographic cosine similarity exceeding 0.85 confirms spatial pattern correspondence across global and regional scales. Related query overlap analysis reveals substantial intersection in verification-focused semantic spaces. Behavioral signature assessment confirms parallel user journey patterns, session characteristics, and professional verification behaviors. Volatility profile comparison demonstrates comparable variance structures, clustering phenomena, and mean reversion properties.

For researchers, the findings advance methodologies for multi-dimensional similarity assessment of search pattern data, demonstrating that comprehensive structural analysis requires examining temporal, spatial, semantic, and behavioral dimensions simultaneously rather than relying on single metrics. For fraud detection systems, recognizing structural similarity enables more sophisticated monitoring strategies utilizing both terms as complementary surveillance indicators with cross-validation capabilities.

For educational institutions, understanding deep structural similarity between institutional searches and fraud-related searches necessitates integrated reputation management strategies addressing both excellence narratives and authentication concerns. Investment in verification infrastructure, proactive incident communication, and transparent credentialing systems become strategic priorities for managing the coupled search behaviors that similarity analysis reveals.

Future research with complete Google Trends data access could quantitatively validate similarity assessments presented through methodological frameworks, enabling precise numerical similarity scores across all examined dimensions. Integration with search log data, longitudinal tracking studies, cross-platform analyses, and experimental validations would further advance understanding of how institutional prestige, fraud awareness, and verification behaviors intertwine in digital information ecosystems.

The similarity between IIT and Fake search patterns is not superficial correlation but reflects fundamental behavioral and structural alignment driven by common underlying processes in educational fraud dynamics and public information-seeking patterns. Recognizing this deep structural similarity provides essential foundation for developing effective institutional strategies, fraud prevention systems, and digital literacy interventions addressing the coupled concerns about educational credential authenticity in an era of increasing verification complexity.

References

[1] GeeksforGeeks. Similarity Search for Time-Series Data. July 2025. Retrieved from: `https://www.geeksforgeeks.org/similarity-search-for-time-series-data/`

[2] Mishra A. Time Series Similarity Using Dynamic Time Warping - Explained. *Walmart Global Tech Blog, Medium.* December 2020. Retrieved from: `https://medium.com/walmartglobaltech/time-series-similarity-using-dynamic-time-warping-explained-9d09119e48ec`

[3] KX Systems. Pattern Matching with Temporal Similarity Search. July 2025. Retrieved from: `https://kx.com/blog/pattern-matching-with-temporal-similarity-search/`

[4] Bader A. How can we quantify similarity between time series? *Gorilla Tech Blog.* February 2023. Retrieved from: `https://tech.gorilla.co/how-can-we-quantify-similarity-between-time-series-ed1d0b633ca0`

[5] Gellér Z, et al. Investigation of Time Series Representations and Similarity Measures for Structural Damage Pattern Recognition. *PMC.* Retrieved from: `https://pmc.ncbi.nlm.nih.gov/articles/PMC3804407/`

[6] Ryaboy M. Classifying Technical Analysis Patterns Using Time-Series Similarity Search. *KX Systems, Medium.* June 2025. Retrieved from: `https://medium.com/kx-systems/classifying-technical-analysis-patterns-using-time-series-similarity-search-cf95282da2a`

[7] Fernandes R. Ultimate Guide to Seasonal Breakdown with Google Trends Data with Code. *Medium*. December 2023. Retrieved from: https://medium.com/@fernandes.manager/ultimate-guide-to-seasonal-breakdown-with-google-trends-data-with-code-7a6dc8db3c7b

[8] Google. FAQ about Google Trends data - Trends Help. Retrieved from: https://support.google.com/trends/answer/4365533?hl=en

[9] HOP Online. Seasonality in Education: Strategies for Off-Peak Lead Generation. May 2025. Retrieved from: https://www.hop.online/blog/seasonality-in-education-strategies-for-off-peak-lead-generation

[10] Collegedunia. IIT Bombay Placement 2025: Highest Package, Average Package, Top Recruiters & Trends. August 2025. Retrieved from: https://collegedunia.com/university/25703-iit-bombay-indian-institute-of-technology-iitb-mumbai/placement

[11] PaymentsJournal. Seasons of Fraud: How Fraud Patterns Shift Throughout the Year. August 2024. Retrieved from: https://www.paymentsjournal.com/seasons-of-fraud-how-fraud-patterns-shift-throughout-the-year/

[12] Deccan Chronicle. Fake PhD Student Busted After 2 Weeks Inside IIT-Bombay. June 2024. Retrieved from: https://www.deccanchronicle.com/nation/fake-phd-student-busted-after-2-weeks-inside-iit-bombay-1888848

[13] Rousseeuw P, et al. Robust Monitoring of Time Series with Application to Fraud Detection. *ScienceDirect*. July 2018. Retrieved from: https://www.sciencedirect.com/science/article/pii/S2452306218300303

[14] Siegler R. Predicting Machine Failure with Time-Series Pattern Matching. *KX Systems, Medium*. June 2025. Retrieved from: https://medium.com/kx-systems/time-series-similarity-search-for-iot-sensor-failure-detection-6573de6c55e4

[15] Google News Initiative. Google Trends: Understanding the data. Retrieved from: https://newsinitiative.withgoogle.com/resources/trainings/google-trends-understanding-the-data/

[16] ScienceDirect. Information-Seeking Behavior - an overview. Retrieved from: https://www.sciencedirect.com/topics/social-sciences/information-seeking-behavior

[17] Semrush. How to Use Google Trends for SEO in 2025. July 2025. Retrieved from: https://www.semrush.com/blog/google-trends/

The End