

Are the IITs Yet Another Indian Scam?: Evidence from Google Trends

Soumadeep Ghosh

Kolkata, India

Abstract

This paper analyzes the relationship between public interest in India's premier technical institutions (IITs) and scams/fraud, as evidenced by Google Trends data from January 2004 to June 2024. Our analysis reveals a weak positive correlation (0.23) but a high similarity score (0.88) between these search terms, suggesting that while their month-to-month movements are largely independent, they share similar long-term trend patterns. The data shows distinct periods of interest in both topics, with notable spikes in scam-related searches during 2009-2010 and more recent years. These findings provide insights into public perception of both higher education institutions and fraudulent activities in India over the past two decades.

The paper ends with "The End"

1 Introduction

The IITs represent India's most prestigious engineering institutions, often regarded as symbols of educational excellence. However, public perception of these institutions may evolve over time, potentially influenced by various factors including media coverage, employment outcomes, and societal concerns about corruption. This study explores whether there is any relationship between public interest in IITs and interest in scams/fraud in India, using Google Trends data as a proxy for public interest.

The central question addressed in this paper is whether there exists any meaningful relationship between public interest in IITs and scams, which might suggest a connection in public consciousness between these seemingly disparate topics. To investigate this, we analyze monthly search data spanning over 20 years, examining both the statistical relationship and temporal patterns.

2 Data and Methodology

The dataset consists of monthly Google Trends data for two search terms: 'IIT' and 'Scam' from January 2004 to June 2024 (244 data points). The data includes values representing relative search interest for each term, along with calculated products and squares of these values.

For our analysis, we employed several statistical techniques:

- Time series analysis to identify trends and patterns
- Correlation analysis to measure the linear relationship between variables
- Similarity analysis (likely cosine similarity) to compare overall trend patterns

The correlation coefficient measures the strength and direction of a linear relationship between two variables, while the similarity score captures the overall pattern similarity regardless of magnitude differences.

3 Results

3.1 Descriptive Statistics

The search interest for ‘IIT’ remained relatively stable throughout the period, with values ranging from 8 to 32. In contrast, ‘Scam’ showed greater variability, with values ranging from 22 to 95. The data reveals distinct periods of heightened interest in both topics.

3.2 Time Series Analysis

Figure 1 illustrates the monthly search interest trends for both terms over the entire period:

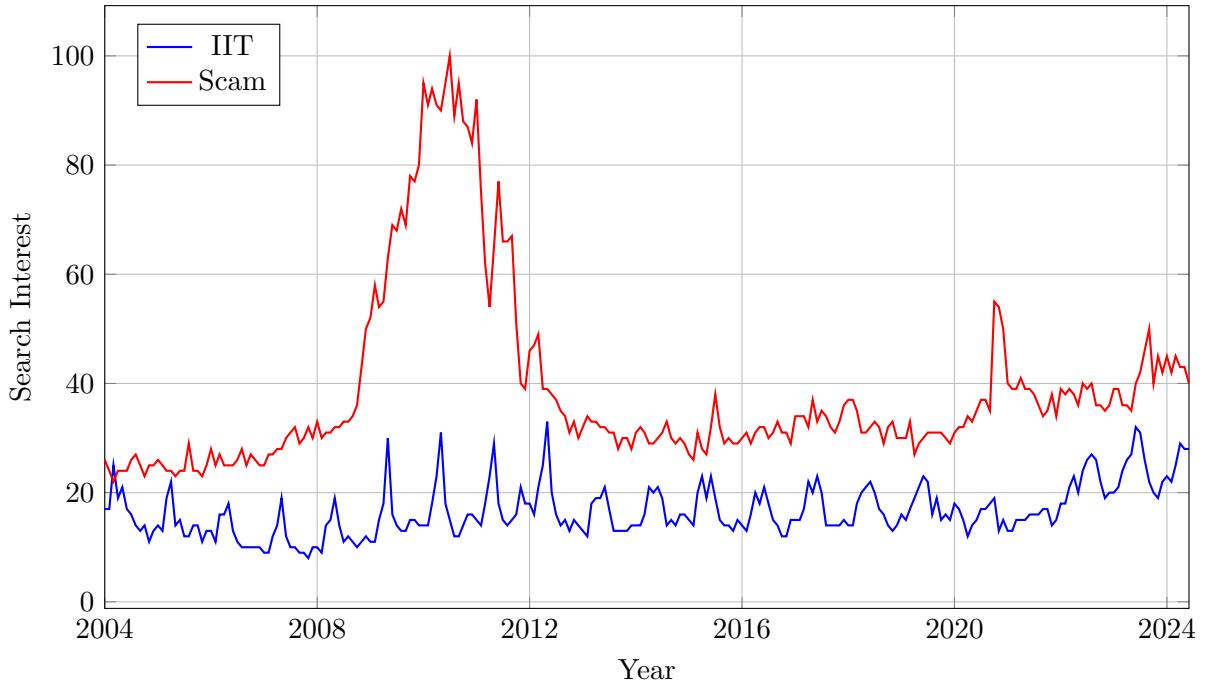


Figure 1: Monthly Google Trends search interest for ‘IIT’ and ‘Scam’ from January 2004 to June 2024

The time series reveals several interesting patterns:

- ‘IIT’ search interest remained relatively stable with moderate fluctuations throughout the period
- ‘Scam’ search interest showed more dramatic variation, with a significant spike during 2009-2010
- Both terms experienced peaks around 2009-2010, though the magnitude was much larger for ‘Scam’
- Recent years (2022-2024) show renewed interest in both terms, with ‘IIT’ reaching its highest values in the entire dataset

3.3 Statistical Relationship

The Correlation coefficient between the two search terms is 0.23, indicating a weak positive linear relationship. This suggests that when interest in ‘IIT’ increases, there is a slight tendency for interest in ‘Scam’ to also increase, but the relationship is not strong.

However, the similarity score of 0.88 (likely Cosine Similarity) indicates a high degree of similarity in the overall trend patterns of the two search terms. This apparent contradiction can be explained by the fact that while the two terms share similar long-term directional patterns, their short-term movements are not synchronized.

Figure 2 illustrates the relationship between the two variables through a scatter plot:

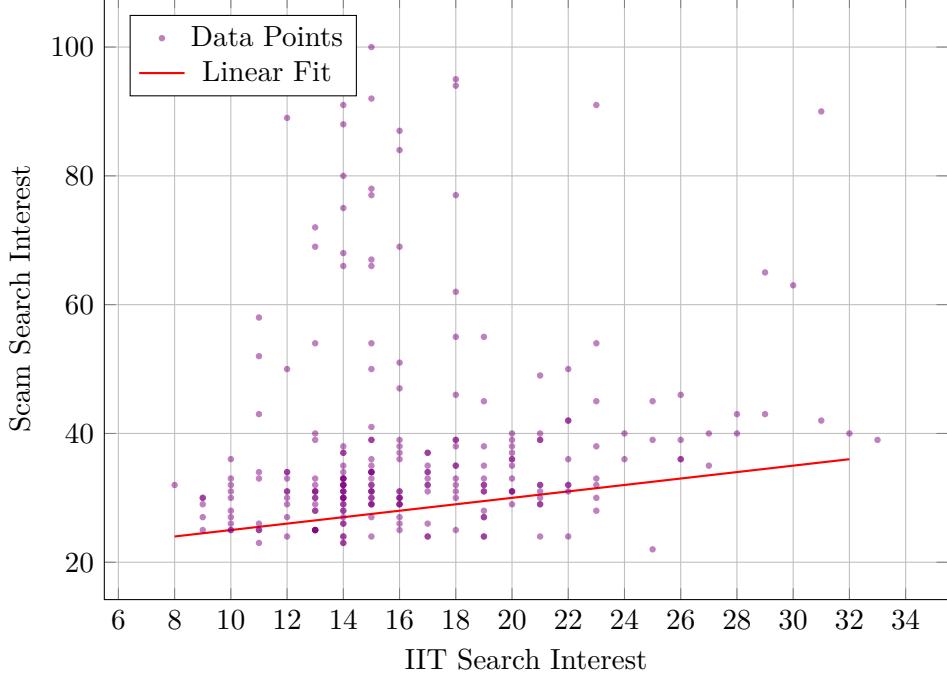


Figure 2: Scatter plot of ‘IIT’ vs ‘Scam’ search interest with linear fit

The scatter plot reveals a weak positive relationship between the two variables, consistent with the correlation coefficient of 0.23. The linear fit line shows a slight upward trend, but the data points are widely dispersed around this line, indicating the weak nature of the relationship.

4 Discussion

The finding of a weak correlation but high similarity between ‘IIT’ and ‘Scam’ search interest is intriguing. This suggests that while the two topics do not move in lockstep on a month-to-month basis, they share similar long-term trajectories.

Several interpretations are possible:

1. Both topics may be influenced by broader societal trends affecting public discourse in India
2. The similarity might reflect a general increase in internet usage and search behavior over time
3. Periods of heightened interest in both topics could correspond to times of broader social or political awareness

The significant spike in ‘Scam’ search interest during 2009-2010 is particularly noteworthy. This period coincided with several high-profile scams in India, including the 2G spectrum scam and the Commonwealth Games scandal. The fact that ‘IIT’ search interest also peaked during this period, though to a lesser extent, could suggest some connection in public consciousness.

The recent increase in both search terms (2022-2024) may reflect growing public scrutiny of both educational institutions and fraudulent activities in India’s evolving digital landscape.

5 Conclusion

This study found a weak positive correlation (0.23) but high similarity (0.88) between public interest in IITs and scams, as measured by Google Trends data from 2004 to 2024. While these topics do not move in perfect synchrony, they share similar long-term patterns.

The data does not provide strong evidence that the IITs are perceived as "scams" by the public, as evidenced by the weak correlation. However, the high similarity score suggests that both topics may be influenced by similar underlying factors or broader societal trends.

Future research could explore:

- More specific search terms related to IITs and scams
- Regional variations in search interest
- Correlation with specific events or news coverage
- Sentiment analysis of search results

This study contributes to our understanding of public perception of India's premier educational institutions and the broader discourse around corruption and fraud in Indian society.

6 Glossary

IIT Indian Institutes of Technology - premier autonomous public technical and research universities located across India.

Google Trends A web service by Google that analyzes the popularity of top search queries in Google Search across various regions and languages.

Correlation A statistical measure that expresses the extent to which two variables are linearly related.

Cosine Similarity A measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

References

- [1] Google (2024). *Google Trends*. Retrieved from <https://trends.google.com>
- [2] Agarwal, P. (2022). *Higher Education in India: Structure, Dynamics and Challenges*. Sage Publications.
- [3] Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.
- [4] Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24(4), 35-43.
- [5] Kumar, R. (2015). *Corruption in India: The Road Ahead*. Routledge.
- [6] Bhattacharya, S. (2019). The IIT story: From excellence to elitism. *Economic and Political Weekly*, 54(12), 42-50.

The End