Digital Camouflage: Investigating Coordinated Inauthentic Activity in Indian Army Social Media Networks

Media Forensic hub

Abstract— This study presents a computational framework for detecting coordinated inauthentic behavior on social media platforms, focusing on Twitter accounts associated with the Indian Army. By integrating linguistic analysis, behavioral pattern recognition, and machine learning techniques, the framework identifies synchronized account activities through metrics such as posting schedules, media sharing patterns, and engagement strategies. Key methodological components include data preprocessing, network analysis, and machine learning classification. Empirical validation on 769 Twitter accounts demonstrates the framework's effectiveness in uncovering potential centralized account management practices, contributing to misinformation detection and online platform integrity maintenance.

Index Terms— Social media, spam, bot behaviour

1 Introduction

The increasing reliance on social media platforms for communication and influence has raised concerns about the authenticity of online discourse, especially in contexts involving national security. The Indian Army's official and unofficial presence on Twitter represents a core element of its public relations strategy, aimed at engaging with both domestic and international audiences. However, this presence also creates opportunities for coordinated inauthentic behavior, which can be used to manipulate public perception through the appearance of consensus or artificially amplified narratives.

By investigating these patterns, this research seeks to enhance understanding of digital coordination within military-affiliated networks. Findings from this study contribute to media forensics by shedding light on potentially centralized account management practices, supporting efforts to maintain the integrity of online platforms and strengthen defenses against coordinated inauthentic activity in matters of national security.

2 Data set

The dataset used in this study comprises 769 Twitter accounts associated with the Indian Army, which were highly active between August and October 2023. These accounts were selected through a rigorous process to identify patterns indicative of coordinated inauthentic activity.

2.1 Data Collection

Selection criteria were designed to focus on accounts that exhibited behavioral similarities, suggesting possible centralized control or management by the same individual or group. The selection criteria are outlined as follows:

- Sharing of Specific Media Items: Accounts that consistently shared identical or closely related images, videos, or other media content about the Indian Army were included in the dataset. This criterion helped identify accounts involved in amplifying specific narratives or media, a common characteristic of coordinated campaigns.

- Synchronized Posting Timings: Accounts that demonstrated highly coordinated posting schedules such as posting at nearly identical times or with minimal intervals between posts were flagged for further analysis. This pattern of synchronized activity suggests centralized control or automated scheduling.

- Client Usage Patterns: Accounts with implausibly similar patterns in the Twitter clients (applications or devices) used for posting were identified as potentially coordinated. Consistency in client usage among multiple accounts is often indicative of single-user control or automated processes.

- Exclusion Criteria: Accounts identified through manual inspection as idiosyncratic or genuinely individual, based on unique posting behaviors or personalized content, were excluded. This step was essential to ensure the dataset was focused on accounts exhibiting characteristics of coordinated activity.

2.2 Data Cleaning

The data cleaning process was a critical step in preparing the dataset for analysis. Initially, the study involved a comprehensive examination of a CSV file containing 769 Twitter accounts. The primary goals of this phase were to organize the data effectively and ensure its relevance to the study's focus on inauthentic coordinated behavior in 2023.

- User-Level Organization The accounts were systematically arranged into user-level JSON files. This restructuring facilitated easier access to individual user data and enhanced the overall organization of the dataset. By transforming the data into a JSON format, it allowed for more efficient data manipulation and analysis in subsequent stages.

- Filtering Inactive Users During the cleaning process, accounts that were not active in 2023 were identified and removed from the dataset. This filtering was essential to focus the analysis on relevant accounts and ensure that the study concentrated solely on the activity occurring within the specified timeframe. As a result of this filtering, the dataset was reduced from 769 to 602 active accounts, significantly narrowing the scope to those contributing to the observed coordinated behavior.

3 Challenges

Investigating coordinated inauthentic behavior across social media accounts introduces a distinctive set of challenges. The analysis of 769 Twitter accounts associated with the Indian Army required overcoming various hurdles to accurately identify potential coordination. The key challenges encountered include:

- Content Overlap Among Accounts: A significant number of accounts in the dataset exhibited content that was remarkably similar, primarily revolving around themes related to the Indian Army. This overlap made it difficult to differentiate between coordinated inauthentic activities and genuine independent expressions of interest. Developing sophisticated methodologies to discern between intentional orchestration and natural content convergence was essential for accurate analysis.

- Ambiguity in Unique User Identification: The initial stages of the study lacked clarity regarding the number of distinct users managing the accounts. This uncertainty posed difficulties in accurately assessing the extent of coordination. Consequently, an adaptive approach to clustering and user identification was necessary, relying on behavioral patterns as primary indicators of potential consolidation under single-user management.

- Differentiating Coordinated Actions from Algorithmic Effects: Similarities in posting behaviors, timing, or shared content might arise from Twitter's algorithms rather than intentional coordination. For instance, the platform may influence content visibility or user engagement timing, leading to unintentional synchronization. Distinguishing between algorithmic effects and deliberate coordination required a thoughtful design of behavioral indicators and robust validation methods.

- Scalability of Analytical Techniques: The analysis of 769 accounts using complex and multifaceted methodologies imposes considerable computational demands. Implementing advanced clustering, temporal analyses, and behavioral segmentation necessitates significant computational resources and optimized algorithms to achieve scalability and efficiency. Striking a balance between computational costs and analytical thoroughness was crucial to conducting a comprehensive investigation within manageable resource limitations.

These challenges underscore the intricacies involved in recognizing coordinated inauthentic behavior within a sizable dataset of social media accounts. Effectively addressing these issues was vital for ensuring the reliability and validity of the findings presented in this study.

4 Methodology:

This study examines coordinated inauthentic behavior on social media by analyzing the activity patterns and shared content features of 602 Twitter accounts associated with the Indian Army. The primary aim is to identify accounts potentially operated by the same entity or by highly synchronized groups. The methodology is structured into four main steps: **Timeline Activity Analysis, Vector Representation of Account Activity, Hard Matching for Automated Account Detection, and Similarity Analysis for Potentially Linked Accounts**.

4.1 Timeline Activity Analysis:

To ensure relevance and remove inactive or outdated accounts, the dataset was refined by focusing on recent activity within the year 2023.

Using the Behavioral Logic for Online Classification (BLOC) framework, we analyzed each account's activity exclusively within 2023, allowing precise tracking of recent behavior patterns.In adapting the BLOC (Behavioral Language Online Characterization) framework for our analysis, we leveraged the existing implementation available on GitHub and tailored it to process static Twitter data. Originally designed for real-time data streams, we modified the tool to operate on archived datasets, enabling retrospective analysis of social media behaviour. This adaptation involved restructuring the codebase to accommodate batch processing of large-scale Twitter datasets, focusing on feature extraction from historical tweet records rather than live feed interactions. By repurposing the BLOC tool for offline analysis, we were able to apply its behavioral language characterization capabilities to study past patterns of social media activity and classify users based on their behaviour profiles.Accounts without significant activity in 2023 were excluded, reducing the dataset from 769 to 602 accounts.

In Figure1, we have explained the workflow of our methodology.

4.2 Vector Representation of Activity

Each account's daily activity in 2023 was encoded as a vector to enable easy comparison. Each account was represented as a daily tweet vector over a 364-day period, with each element representing the daily tweet count.These vectors were stored in a structured format to facilitate efficient comparative analysis.Representing accounts as vectors allows for effective comparison of posting patterns, enabling the detection of similarities indicative of coordinated behavior.

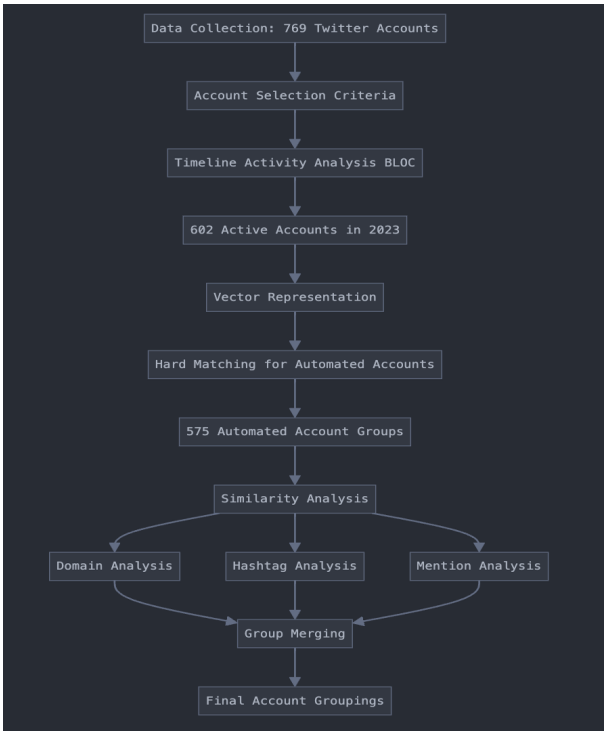

Figure 1: Workflow of the Methodology

4.3 Automated Account Detection

To identify accounts exhibiting identical or near-identical automated posting patterns, we applied a hard matching algorithm. This algorithm detected accounts with highly similar or identical activity vectors, signaling potential automation or highly synchronized human activity. Accounts with matching vectors were grouped, with each group representing a potential automated cluster. This step yielded 575 groups, marking these accounts as potentially automated or reflecting tightly coordinated posting schedules. This initial grouping isolated accounts likely involved in automation or structured coordination, forming a foundation for further analysis.

4.4 Similarity Analysis for Potentially Linked Accounts

To detect more subtle coordination patterns, a similarity analysis was conducted based on shared content features. This analysis aims to capture accounts potentially operated by the same entity or organized group by evaluating shared items in their posts.For each identified group, we analyzed shared domains (Frequently shared URLs or sources), common hashtags(Recurrent hashtags used by group members),mentioned accounts(Other accounts frequently mentioned by group members).

Similarity Calculation: The similarity between two

groups was calculated separately for domains, hashtags, and mentions using set-based similarity. Each feature was scored by dividing the count of common items by the size of the larger set:

$$similarity = \frac{|commonitems|}{\max(|set1|, |set2|)}$$

A combined similarity score was obtained by averaging the individual scores for hashtags, domains, and mentions:

$$score = \frac{hashtagsimilarity + domainsimilarity + mentionsim}{3}$$

4.5 Threshold Based Merging

If the combined similarity between two groups exceeded a threshold (initially set at 0.85), the groups were merged. The smaller group was typically merged into the larger one, or the group with the lower ID if sizes were equal.To capture increasingly subtle patterns, the similarity threshold was gradually reduced from 95% down to 50% in increments of 5% across iterations.This approach enabled the identification of coordination patterns that extended beyond posting behavior, capturing cross-group connections based on shared content.

Figure2 illustrates the results of varying thresholds, highlighting how the number of unique groups changes as the threshold decreases. Post-grouping validation further ensures the appropriateness of these thresholds. By analyzing average hourly activity patterns within each identified group, we confirm that the merging results are consistent with observed behavioral synchronization. Groups that exhibit tightly aligned activity provide strong evidence that the chosen threshold accurately captures underlying coordination patterns.

5 Observation

A key observation in our analysis is the consistent alignment between daily-level and hourly-level activity across accounts. This correlation has been repeatedly observed throughout the dataset, providing a reliable pattern of behavior. We leveraged this temporal consistency as a cornerstone in the development of our methodology, using it to refine account groupings and improve the overall robustness of our analysis.

In Figure3, we present the average hourly activity of seven accounts that have been grouped based on their average daily tweet activity over the course of the year. Each line in the visualization represents the activity pattern of a different user. Ideally, we would expect to see seven distinct colored lines, each corresponding to one user's average hourly activity. However, due to the users

```
Threshold: 95%, Unique Groups: 429
Threshold: 90%, Unique Groups: 371
Threshold: 85%, Unique Groups: 219
Threshold: 80%, Unique Groups: 131
Threshold: 75%, Unique Groups: 74
Threshold: 70%, Unique Groups: 51
Threshold: 65%, Unique Groups: 37
Threshold: 60%, Unique Groups: 27
Threshold: 55%, Unique Groups: 21
Threshold: 50%, Unique Groups: 14
Threshold: 45%, Unique Groups: 11
Threshold: 40%, Unique Groups: 7
Threshold: 35%, Unique Groups: 6
Threshold: 30%, Unique Groups: 5
Threshold: 25%, Unique Groups: 5
Threshold: 20%, Unique Groups: 5
Threshold: 15%, Unique Groups: 5
```

Figure 2: Groups after Merging

within this group exhibiting identical average hourly activity patterns, only a single line is visible, which represents the uniformity in the average hourly activity of these grouped users.

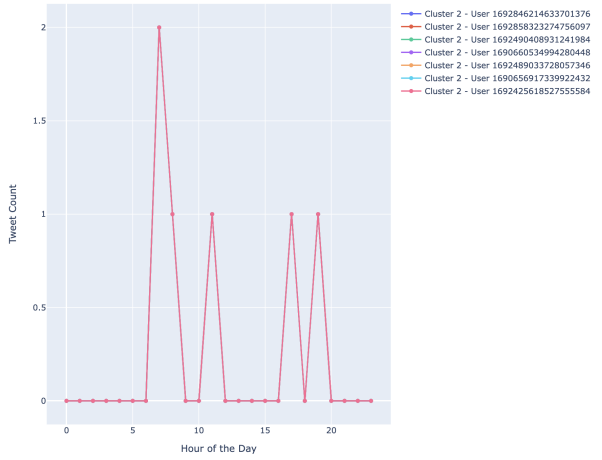

Figure 3: Average hourly tweet activity

5.1 Evaluation

The evaluation of the proposed methodology focuses on its effectiveness in identifying coordinated inauthentic behavior among the 602 Twitter accounts associated with the Indian Army. Given the unsupervised nature of this research, the absence of a known ground truth presents

unique challenges in assessing the performance and validity of the findings. This section outlines the evaluation approach, insights gained, and considerations for future research.

Correlation Matrix Analysis: As the first step in evaluating the groups, a correlation matrix was generated from the automated users file. This matrix examined the similarities in hashtags, domains, and mentions among users within a particular group.

Domain Similarity: The matrix highlighted the commonality of shared URLs or sources among users, showcasing the degree of overlap in the content they engaged with. Figure4 displays a matrix where user IDs of individuals within the same cluster are organized along both axes. The heatmap utilizes shades of red to highlight areas of maximum domain similarity, indicating the strongest connections among users within a cluster.

Hashtag Similarity: The correlation coefficients indicated how frequently group members used the same hashtags in their posts. Figure5 displays a matrix where user IDs of individuals within the same cluster are organized along both axes. The heatmap utilizes shades of red to highlight areas of maximum hashtag similarity, indicating the strongest connections among users within a cluster.

Mentions Similarity: Analyzing the correlation in mentioned accounts revealed connections between users in terms of interactions and engagements. Figure6 displays a matrix where user IDs of individuals within the same cluster are organized along both axes. The heatmap utilizes shades of red to highlight areas of maximum mentions similarity, indicating the strongest connections among users within a cluster.

Evaluation of Group Stability Over Time:

To validate the robustness of our account groupings, we extended our analysis one year into the future. We observed that accounts within each group continued to exhibit highly synchronized tweet activity, with multiple accounts posting at the same precise time every hour. Furthermore, several accounts in the same groups were suspended nearly simultaneously, further supporting our hypothesis. These results provide compelling evidence that these accounts are likely operated by the same individual or group, strengthening the case for coordinated inauthentic behavior.

6 Results and Conclusion

Our analysis provides strong evidence of coordinated inauthentic behavior among the studied Twitter accounts. By focusing on temporal activity patterns, we were able

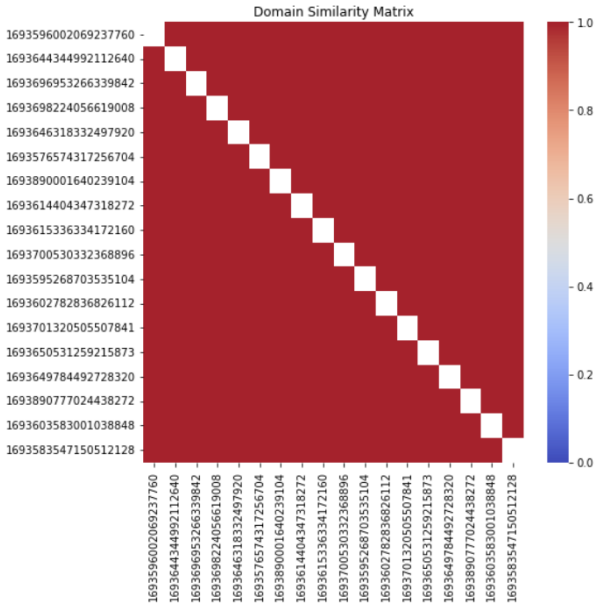

Figure 4: Domain Similarity matrix

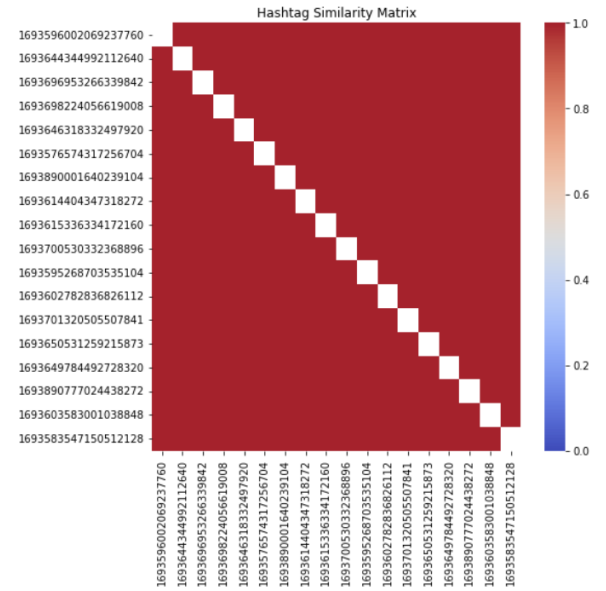

Figure 5: Hashtag Similarity matrix

to consistently group accounts that exhibited synchronized tweeting behavior. This synchronization, along with the simultaneous suspension of several accounts, strongly indicates that these accounts are likely operated by a small number of users or a single entity. Over the course of one year, we observed a consistent pattern in which users appeared to work in pods of 2-3 accounts, with their activity maintaining close alignment throughout the study period. This consistent activity suggests a

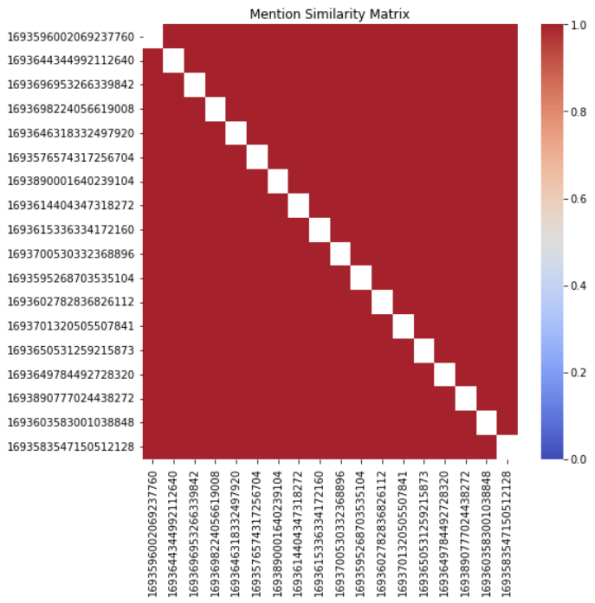

Figure 6: Mention Similarity matrix

deliberate strategy aimed at amplifying their online presence.

These findings are in line with the observations made in the report "My Heart Returns to Kashmir," which served as a reference for this study. The report highlights the prevalence of coordinated behavior among online entities and provided a foundational understanding of the phenomena we observed in our dataset. By analyzing the number of accounts controlled by a single entity or group, we identified a pattern that could inform the establishment of more refined thresholds for detecting coordinated inauthentic activity in social media networks.

The threshold based merging analysis provides a flexible framework to tailor the similarity threshold according to the user's objectives and the expected number of unique groups operated by the same entity. For example, if a higher granularity is required, a threshold near 95% may be preferred, resulting in a larger number of unique groups. Conversely, reducing the threshold to 50% or below yields fewer, more consolidated groups.

To ensure the appropriateness of the chosen threshold, post-grouping analysis can be conducted. By examining the average hourly activity patterns for each identified group, we validate that the merging results align with observed behavioral consistency. For instance, groups that exhibit synchronized activity at a granular level provide strong evidence that the threshold accurately captures underlying coordination patterns.This threshold-based methodology, combined with post-merging validation, offers a robust approach to uncovering and analyzing coordinated user behavior.

Our results not only validate the methodology we employed but also offer valuable insights into how such behavior evolves over time. The synchronized activity observed within the groups—despite the accounts being operated separately—suggests that the individuals or groups behind these accounts are working systematically to influence online discourse. These insights contribute to a deeper understanding of how coordinated inauthentic behavior is structured and how it can be better detected in future studies.