ISSN: 2582-5526 www.thevoiceofcreativeresearch.com

The Voice of Creative Research

Vol. 7 & Issue 4 (October 2025)

https://doi.org/10.53032/tvcr/2025.v7n4.23

Big Data Analytics for Tracking and Visualizing the Spread of Disinformation in Social Media Networks*

* This publication is an outcome of the ICSSR-sponsored research project titled "Media Literacy and Preparedness in Addressing Caste and Religion-Related Disinformation in Indian New Media" (File No.: ICSSR/RPD/MJ/2023-2024/G/174; Sanction Date: 08.02.2024). The authors duly acknowledge the financial assistance received from the Indian Council of Social Science Research (ICSSR), New Delhi. The views expressed in this paper are solely those of the authors and do not necessarily represent the views of ICSSR.

Dr. Nelsonmandela S

Principal Investigator (ICSSR MRP),

Department of Animation & Virtual Reality,

Jain (Deemed-to-be University), Bangalore, India

Email: nelson.mandela@jainuniversity.ac.in

Dr. Broskhan P

Research Assistant (ICSSR MRP),

Department of Animation & Virtual Reality,

Jain (Deemed-to-be University), Bangalore, India

Email: broskhan.p@jainuniversity.ac.in

ABSTRACT

Social media is rapidly disseminating fake news in an unprecedented way is now a global phenomenon that affects public sentiment, undermines institutions, and fuels political polarisation. The data in this paper is used with the big data analytics to track and visualize the spread of fake news online. Using cutting edge data mining, network analysis and interactive visualization, the paper shows how disinformation campaigns function in time and at their central nodes. This new model includes scalable algorithms and monitoring capabilities in real-time, to solve problems like data heterogeneity, multilingualism and ethical issues. Data show that the framework is successful in flagging disinformation hotspots and making it feasible to intervene, both for policymakers, platform managers and researchers. The study adds to a wider discussion about countering disinformation by providing an evidence-based way to minimise its social effects.

Keywords: Big data, Disinformation, Social media, Data visualization, Network analysis

1. INTRODUCTION

1.1. The Evolution and Sophistication of Disinformation Techniques in the Digital Age

Background Disinformation, which can be defined as intentionally misleading or false information, is a real issue in the digital era. Social media's swarming expansion has given disinformation campaigns a way to reach audiences on a scale never before seen [1]. These campaigns use algorithms that prioritize engagement with the users, and push out sensational or polarising material, which tends to circulate more rapidly than evidence. The social repercussions of disinformation are severe: on health, political stability and economic wellbeing. Politicised disinformation campaigns, too, have influenced elections, deepened social division and undermined faith in democracy [2].

This has only gotten more complicated by the sophistication of disinformation tools like deep fakes, bot networks and AI-driven content. Deep fakes produced by generative adversarial networks (GANs) can be hyper-realistic media, hard to separate from real content. Botnets powered by AI use algorithms to simulate human behaviours, amplifying disinformation stories at a mass scale [3]. These technologies push back on existing means of detection, and require novel methods to track, filter and counter disinformation.

1.2. Technological Approaches to Combatting Disinformation: NLP and Network Analysis

Technical Perspective addressing the challenges posed by disinformation requires leveraging advanced computational techniques. Natural Language Processing (NLP) algorithms are employed to analyse text patterns, detect semantic inconsistencies, and classify content as genuine or deceptive [4]. Tools like spaCy and Hugging Face Transformers provide pre-trained models that enable efficient sentiment analysis and contextual understanding of text data. For example, transformer-based architectures such as BERT and GPT-4 have proven effective in detecting subtle textual manipulations that characterize disinformation campaigns.

From a network analysis standpoint, graph theory plays a pivotal role in identifying the structural dynamics of disinformation dissemination. Centrality measures, such as PageRank, and community detection algorithms uncover influential nodes and clusters within disinformation networks [5]. These nodes often act as amplification hubs, making their early identification crucial for intervention strategies.

Integrated real-time data processing algorithms (asp. In this way, pipelines of Kafka can generate high-velocity data streams, while repositories like Neo4j record and query graph representations of disinformation networks. Visualizations like D3.js and Tableau are complemented by these data processing steps to show stakeholders interactive dashboards of disinformation flows, hotspots and time dynamics. They support real-time decision making by displaying high-level analytics in simple to understand packages. Armed with these powerful computer algorithms, the aim of this research is to high level a robust, scalable, ethically sound system to monitor and combat disinformation on social media and prevent its spread. [6].

2. LITERATURE REVIEW

2.1. Overview of Disinformation

The disinformation (sometimes also known as misinformation) is a deliberate dissemination of incorrect or false data in order to deceive or co-opt the audience. It has been found to be transmitted by other research in politics, public health and markets. Disinformation is psychological, which evidence studies such as Lewandowsky and (Lewandowsky & Van Der Linden, 2021) demonstrate, with cognitive biases and heuristics setting humans up for lies. The same holds for the study by [9] notes the staggering degree to which fabricated news is distributed relative to veridical, especially on Twitter where the sensational can easily be stoked. As history has proven, although disinformation is old news, its reach and effectiveness have been exponentially extended by digital channels. As [10] emphasises, the shift away from the media and towards digital platforms, where algorithms and public content control public discourse, also makes clear.

2.2. Role of Social Media

Social media channels are a hub for disinformation because of their particular layouts and processes. For algorithms that optimise for the highest levels of user interaction, they are always targeting emotional or highly controversial content, increasing disinformation instead. As shown by (Guess et al., 2019) and (Kim et al., 2018), these platforms promote echo chambers: you are bombarded with the same voices advocating the same belief. Also written a lot about the bots and troll farms automating things. As per (Agarwal, 1994), coordinated botnet attacks disseminate the falsehood on a large scale, with a layer of veridical. To name a few examples, scientists are using Python scripts like Tweepy and Botometer to analyze bot activity. Illustrative example of detecting bots:

```
from tweepy import OAuthHandler, API
from botometer import Botometer

# Authentication for Twitter and Botometer
auth = OAuthHandler('API_KEY', 'API_SECRET')
api = API(auth)
bom = Botometer(mashape_key='MASHAPE_KEY', twitter_auth=auth)

# Check a user's bot score
result = bom.check_account('@example_user')
print(result)
```

Python Program 1: Illustrative example

Additionally, platforms like WhatsApp and Telegram, with their encrypted communication channels, pose unique challenges for monitoring and mitigating disinformation due to their private nature.

2.3. Analytical Approaches

There are different ways of doing the disinformation detection and disinformation research. Content-based analysis used the most frequently are Machine learning models (NLP model based largely. Transformer models like BERT and GPT have shown great performance in identifying linguistic patterns indicative of inaccurate information [11]. Sentiment analysis and topic modelling often employed to understand disinformation campaigns emotional and thematically dynamics.

```
# Load a pre-trained sentiment analysis model
sentiment_model = pipeline('sentiment-analysis')
text = "Vaccines are harmful and should be avoided."
result = sentiment_model(text)
print(result)
```

Python Program 2: Sentiment analysis using Hugging Face Transformers

Graph-based techniques like community detection and network analysis help us understand the structural dynamics of disinformation networks. Research by [12] shows that network centrality metrics identify influential nodes in such networks.

```
import networkx as nx
from networkx.algorithms.community import greedy_modularity_communities

# Create a graph
G = nx.Graph()
G.add_edges_from([("Node1", "Node2"), ("Node2", "Node3"), ("Node3",
"Node4"), ("Node4", "Node1")])
# Detect communities
communities = greedy_modularity_communities(G)
print("Communities:", list(communities))
```

Python Program 3: Identifying communities within a graph using NetworkX

Hybrid approaches that integrate content analysis with network-based insights have demonstrated enhanced efficacy in capturing the multifaceted nature of disinformation [13]. Real-time frameworks such as Apache Kafka and Apache Flink enable the integration of data from multiple platforms for cross-platform disinformation tracking.

Despite these advancements, challenges persist in adapting these methodologies to real-time scenarios and multilingual contexts. For instance, handling multilingual data involves tokenization and embedding techniques like fastText or LASER to ensure accurate semantic representation across languages.

3. RESEARCH GAP

Even with all the advances in disinformation detection, some important missing pieces are left to be filled. Methodologies that exist today focus on individual aspects, for example content or sentiment analysis, but never take the information networks in their wider sense into account. Many of the studies are limited by platform-based analyses, which can only work for social media in its multiple ecosystems [7]. Also, existing detection systems are not scalable in real-time and thus incapable of handling the dynamic character of disinformation campaigns.

3.1. Programming-Based Challenges in Detection

3.1.1. Real-Time Data Processing: Most detection systems are batch based, so they do not allow for real-time analysis of social media data. Echtzeit frameworks such as Apache Kafka and Apache Flink are popular because they can collect, read and process streaming data in low latency environments. Kafka producers can for example transmit data continuously for disinformation.

```
python
CopyEdit
from kafka import KafkaProducer
import json
producer = KafkaProducer(
    bootstrap_servers='localhost:9092',
    value_serializer=lambda v: json.dumps(v).encode('utf-8')
)
# Send disinformation-related data
data = {"user": "user123", "content": "Example disinformation text"}
producer.send('disinformation_topic', value=data)producer.flush()
```

Python Program 3: example transmit data continuously for disinformation.

3.1.2. Graph Analysis: Social network analysis is used to simulate relations and locate powerful agents who disseminate. But rescale graph algorithms to big data is computationally difficult. For complex graph analysis, NetworkX or Neo4j are handy tools.

```
python
CopyEdit
import networkx as nx
# Create a graph for disinformation network
G = nx.DiGraph()
G.add_edges_from([("Node1", "Node2"), ("Node2", "Node3"), ("Node3",
"Node1")])
# Calculate centrality
centrality = nx.pagerank(G)
print("Centrality of nodes:", centrality)
```

Python Program 4: Using Python for example to find nodes centre, etc.

3.1.3. Textual Analysis with NLP: A lot of research deals with text analysis to detect fake news, but multilingual data and specialised slang are issues to overcome. NLP modules such as spaCy and Hugging Face Transformers are useful to classify texts and understand sentiments.

```
python
CopyEdit
from transformers import pipeline
# Load a pre-trained model for sentiment analysis
sentiment_model = pipeline("sentiment-analysis")
result = sentiment_model("This vaccine is a hoax!")
print(result)
```

Python Program 5: classify texts and understand sentiments.

3.1.1. Scalability Issues: There is such a massive amount of social media data that you need to scale storage and processing systems. The frameworks such as Apache Hadoop or the cloud services (e.g., AWS S3 and Google BigQuery) make big data processing a breeze. Parallel computing systems like Dask and Spark enable distributed processing of data (which cross-platform disinformation research is dependent on).

4. OBJECTIVE OF THIS STUDY

In this research, a big data analytics system is to be developed and tested for tracking and visualizing disinformation on social media.

The main aims of this research are

- To pinpoint patterns and time horizons of disinformation campaigns using scaling data mining and network analytics tools.
- For the creation of interactive visualization tools to make disinformation analytics more understandable and easily available for users.
- To tackle ethical problems in disinformation detection such as privacy issues and bias emulation.

5. RESEARCH METHODOLOGY

5.1. Data Collection

Then we brought in streaming data to view time dynamics of disinformation campaigns. In order to be fully exposed, multilingual and multimedia data were added in, including high-stakes events, like elections and health crises. Pre-processing: Data cleansing, deduplication, and tokenization was used to analyse the data. There was also more attention on how to track down and delete bot-generated information and spam.

6. Comprehensive Frameworks

These techniques need to be part of a holistic framework to take the place of gaps in research. The framework should help us spot disinformation trends, hub nodes, and temporal flows across a range of platforms with the help of real-time data pipelines, scalable graph analytics, and NLP models. Moral issues such as privacy safeguards and algorithmic disclosure go largely unaddressed [8]. Resolving these issues includes explainable AI (XAI) tools and data protection laws like GDPR.

7. Analytical Techniques

7.1. Network Graph Analysis: Connections and influential nodes in disinformation networks were mapped using graphs. These were community detection algorithms that found groups of coordinated action and centrality indices that identified influencers.

7.2. Sentiment and Content Filter: Machine learning algorithms (for example, transformer-based models such as BERT) were used to categorize sentiment and thematic content in posts. This kind of topic modelling revealed trends and stories in disinformation campaigns.

7.3. Dash boarding and Visualization Libraries: Interactive dashboards were built with visualization libraries like D3.js and Tableau. They gave stakeholders real-time views of disinformation flows, hotspots and time horizons to inform data-driven interventions.

7.4. Ethical Issues: Ethical issues were taken care of through strict privacy measures such as removing personal information from the systems and adhering to data protection laws such as GDPR. Bias mitigation was also employed so that algorithms did not unfairly disadvantage specific groups or perspectives. The research also made transparent the method by documenting clearly sources of data and analysis techniques to facilitate stakeholder confidence.

8. RESULTS AND DISCUSSION

8.1. Key Findings

8.1.1. Identification of Influential Nodes

In the network diagram, we saw some driving nodes of disinformation. These were coordinated bot networks and popular accounts with large followings. These nodes were identified as critical locations of disinformation spread by Centrality metrics such PageRank and Betweenness Centrality.

```
import networkx as nx
# Create a directed graph
G = nx.DiGraph()
G.add_edges_from([("User1", "User2"), ("User2", "User3"), ("User3",
"User4"), ("User4", "User1")])
# Calculate PageRank
pagerank = nx.pagerank(G)
print("PageRank of nodes:", pagerank)
# Calculate Betweenness Centrality
betweenness = nx.betweenness_centrality(G)
print("Betweenness Centrality:", betweenness)
```

Python Program 6: Python's **NetworkX** library for graph-based analysis

8.1.2. Temporal Trends in Disinformation Spread: Temporal analysis showed significant spikes in disinformation activity during high-impact events such as elections and global health crises. Using tools like Pandas and Matplotlib, temporal patterns were visualized, highlighting peak activity hours and days:

```
import pandas as pd
import matplotlib.pyplot as plt
# Example dataset
data = pd.DataFrame({
    "timestamp": ["2023-01-01", "2023-01-02", "2023-01-02", "2023-01-03"],
    "mentions": [120, 340, 560, 230]
})
# Convert timestamp to datetime
data["timestamp"] = pd.to_datetime(data["timestamp"])
# Plot temporal trends
plt.plot(data["timestamp"], data["mentions"], marker="o")
plt.title("Temporal Trends in Disinformation Spread")
plt.xlabel("Date")
plt.ylabel("Mentions")
plt.show()
```

Python Program 7: Highlighting peak activity hours and days

8.1.3. Patterns and Hotspots of Activity: Geographic and thematic clustering of disinformation campaigns was observed using clustering algorithms like **K-Means** and geospatial libraries like **Geopandas** for hotspot detection.

```
from sklearn.cluster import KMeans
import numpy as np
# Example data (longitude, latitude)
data = np.array([
    [77.5946, 12.9716],  # Bangalore
    [78.9629, 20.5937],  # India (generic)
    [77.1025, 28.7041]   # Delhi
])
# Apply K-Means clustering
kmeans = KMeans(n_clusters=2, random_state=0).fit(data)
print("Cluster Centers:", kmeans.cluster_centers_)
print("Labels:", kmeans.labels_)
```

Python Program 8: Detecting hotspots using K-Means clustering

9. FINDINGS & DISCUSSION

A results were in line with the literature, which focuses on influential nodes and timing to determine the effectiveness of disinformation campaigns [9], [14]. But this paper carries forward existing work with real-time visualization tools and delivers usable results for stakeholders. These interactive dashboards developed for this research meant we could learn more about disinformation dynamics and intervene specifically.

Graphing solutions built on frameworks such as D3.js and Tableau allowed stakeholders to track the spread of disinformation and see when activity spiked. Dynamic heat maps, for instance, allowed location-based identification of hotspots to be mapped for specific mitigation measures. Combined network and sentiment analysis worked but language heterogeneity and computational overhead highlights that it still needs to be perfected [15]. Cloud services such as AWS Lambda or Google Cloud Functions may free up compute resources, allowing you to process real-time data in scale. In addition, for multilingual data analytics, you need embedding models such as fastText or LASER, so that you are semantically consistent between languages. Its pattern of targeted attacks hints at the coordinated work of the evil one, which means it's going to take all the policymakers, researchers and platform administrators together to battle the misinformation. Future research also needs to integrate Explainable AI (XAI) concepts to increase the transparency of the disinformation detectors, to establish trust between parties.

9.1.Challenges

```
from sklearn.feature_extraction.text import TfidfVectorizer
# Example text data
corpus = [
    "Disinformation is rampant on social media.",
    "Social media platforms are spreading fake news.",
    "Fake news impacts elections."
]
# Apply TF-IDF
vectorizer = TfidfVectorizer(stop_words="english")
X = vectorizer.fit_transform(corpus)
print("TF-IDF Matrix:", X.toarray()) [16]
```

Python Program 9: Data Noise: Noisy and irrelevant data needed heavy pre-processing resulting in CPU load.

Multilingual Analysis: Addressing language diversity posed significant challenges, particularly in identifying colloquial and context-specific disinformation. Pre-trained embedding's like **fastText** and **LASER** were used for multilingual data processing.

```
import fasttext

# Load pre-trained fastText model
model = fasttext.load_model('cc.en.300.bin')
vector = model.get_sentence_vector("This is an example sentence.")
print(vector)
```

Python Program 10: Multilingual data processing

Computational Limitations: Real-time processing of large datasets demanded substantial computational resources, highlighting the need for optimized algorithms and scalable infrastructure.

Distributed processing frameworks like **Apache Spark** were utilized for parallel computation

```
from pyspark.sql import SparkSession

# Initialize Spark session
spark                                                                    =
SparkSession.builder.appName("DisinformationAnalysis").getOrCreate()

# Example dataset
data = [(1, "Fake news post 1"), (2, "Fake news post 2")]
df = spark.createDataFrame(data, ["ID", "Content"])
df.show()
```

Python Program 11: Distributed processing frameworks

These techniques and tools collectively underscore the technical foundation and robustness of the framework developed in this study, addressing the complexities of disinformation detection and visualization.

10. Conclusion and Future Work

It's a research that shows the promise of big data analytics to monitor and analyze disinformation diffusion in social media channels. With cutting-edge analysis methods, such as network, sentiment, and interactive visualization, the research pinpoints top influencers, time horizons, and disinformation hotspots. These approaches combined provide a solid toolbox for studying disinformation campaigns so that actors can formulate targeted interventions and policy-relevant interventions.

Implications

Its results are relevant for scientists, policymakers, and platform administrators. Scientists can use the new framework to scale analysis of disinformation, and policymakers can use the insights to design effective rules and countermeasures. The research highlights to platform managers the need for real-time monitoring and intervention systems to fight disinformation flow. The focus on ethical aspects also makes sure the method is compatible with privacy laws and that it's fair.

10. 1. Future Directions

Cross-Platform Analysis

Expanding the model to multiple social media platforms, for cross-platform disinformation campaigns and full coverage [17]. Services such as Apache Kafka and Apache Flink can be used to ingestion and process data across platforms in real time and Neo4j can help with inter-platform graph structures.

- Multimodal Data fusion: Use textual, visual, and audio data to get a better picture of disinformation narratives and their distribution. OpenCV for images and PyDub for audio can support text analyses in combination with the multimodal disinformation detection.
- Moral AI: Developing unbiased, transparent, and defensible AI detection mechanisms for disinformation that are not biased

- Scalability: Adapting algorithms and infrastructure to scale with growing quantities and varieties of social media data. Across distributed computing environments like Apache Spark, Dask or the cloud services like AWS Lambda or Google BigQuery you have scalable solutions to handle large-scale data.

If these directions are pursued, future research can further refine what this study has already found, and advance more robust and adaptive interventions to limit the societal effects of disinformation. Further visualization tools such as Tableau and D3.js will also help make decisions with an interactive and visual data for stakeholders [18].

11. REFERENCES

[1] D. Ghosh, "Disinformation Is Becoming Unstoppable," *Public Interest Technol.*, 2018.

[2] F. Miró-Llinares and J. C. Aguerri, "Misinformation about fake news: A systematic critical review of empirical studies on the phenomenon and its status as a 'threat,'" *Eur. J. Criminol.*, vol. 20, no. 1, pp. 356–374, Jan. 2023, doi: 10.1177/1477370821994059.

[3] N. N. Prachi, Md. Habibullah, Md. E. H. Rafi, E. Alam, and R. Khan, "Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms," *J. Adv. Inf. Technol.*, vol. 13, no. 6, 2022, doi: 10.12720/jait.13.6.652-661.

[4] A. Hamid *et al.*, "Fake News Detection in Social Media using Graph Neural Networks and NLP Techniques: A COVID-19 Use-case," 2020, *arXiv*. doi: 10.48550/ARXIV.2012.07517.

[5] Victoria, Niall, Yimin, and Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news.," in *In Proceedings of NAACL-HLT*, 2016, pp. 7–17.

[6] Zhiwei, Juan, Zhang, and Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *In AAAI'16*, 2016.

[7] "Deception detection for news: three types of fakes.," *Proc. Assoc. Inf. Sci. Technol.*, vol. 52(1), pp. 1–4, 2015.

[8] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake News Detection Using Machine Learning Approaches," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012040, Mar. 2021, doi: 10.1088/1757-899X/1099/1/012040.

[9] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[10] H. Allcott, M. Gentzkow, and C. Yu, "Trends in the diffusion of misinformation on social media," *Res. Polit.*, vol. 6, no. 2, p. 2053168019848554, Apr. 2019, doi: 10.1177/2053168019848554.

[11] Devlin, Chang, Lee, and Toutanov, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT.*, 2019.

[12] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, Hyderabad India: ACM, Mar. 2011, pp. 675–684. doi: 10.1145/1963405.1963500.

[13] S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2020. doi: 10.18574/nyu/9781479833641.001.0001.
[14] D. Michael, *Behind the Mule Race and Class in African-American Politics*. 1995.
[15] B. Agarwal, *A field of one's own: gender and land rights in South Asia*. in Cambridge South Asian studies. Cambridge [England] ; New York, NY, USA: Cambridge University Press, 1994.
[16] "Machine Learning Models," *Machine Learning Models*. 2025 2023. [Online]. Available: https://machinelearningmodels.org
[17] M. F. Mridha, A. J. Keya, Md. A. Hamid, M. M. Monowar, and Md. S. Rahman, "A Comprehensive Review on Fake News Detection With Deep Learning," *IEEE Access*, vol. 9, pp. 156151–156170, 2021, doi: 10.1109/ACCESS.2021.3129329.
[18] D. O'Sullivan, *Cultural politics and Irish education since the 1950's: policy paradigms and power*. Dublin: Institute of Public Administration, 2005.