Misinformation Cascade Analysis on X (Twitter)

Ahmad Suleiman
University of Illinois at Chicago
Email: asule@uic.edu
UIN: 678457529

Celine Taki
University of Illinois at Chicago
Email: ctaki2@uic.edu
UIN: 661775426

Pradyun Shrestha
University of Illinois at Chicago
Email: pshre3@uic.edu
UIN: 651215081

Niharika Belavadi Shekar
University of Illinois at Chicago
Email: nbela@uic.edu
UIN: 675234184

Andy Mai
University of Illinois at Chicago
Email: amai4@uic.edu
UIN: 674785555

Jeremy Albios
University of Illinois at Chicago
Email: jalbi2@uic.edu
UIN: 663036706

Group Leader: Niharika Belavadi Shekar
Group Name: Fact Checkers

***Abstract*—This project analyzes how misinformation emerges, spreads, and evolves across X (Twitter) by modeling the complete life cycle of false information. We examine how misinformation originates, identify key amplifiers, and track temporal diffusion patterns. Using network analysis techniques including link analysis, community detection, and cascade modeling, we aim to expose the underlying structures and behavioral mechanisms that enable viral spread of misinformation compared to factual content.**

I. Problem and Goal

A. Problem Statement

Our team wants to analyze how misinformation emerges, spreads, and evolves across X (Twitter) by modeling the complete life cycle of a false piece of information. We want to understand how misinformation originates, who initiates it, and which users act as amplifiers through retweets or replies. By examining how information travels through the network over time, we can identify key influencers and the dynamics that drive its spread. This project involves analyzing user interactions to uncover influential nodes, mapping the development of misinformation communities, and tracking temporal diffusion patterns that reveal when and how quickly misinformation cascades expand. Ultimately, our goal is to expose the underlying network structures and behavioral mechanisms that allow misinformation to spread virally compared to factual content.

B. Motivation

This topic is important because misinformation has significant real-world consequences. It can mislead the public during elections, health crises, and emergencies, creating confusion, distrust, and even fearmongering, where false information intentionally stirs fear or panic within communities. The majority of people in our generation use social media, users often tend to believe the first thing they see, which allows misinformation to spread rapidly and shape public opinion. False narratives can alter our perception of people, places, and events, influencing how we interact with the world both socially and economically. Paid bots are often used to amplify the reach and impact of misinformation by pushing propaganda or specific ideologies. This problem affects a wide range of users: young people who are still developing critical thinking skills and older generations who may struggle to recognize AI-generated content, such as deepfakes or videos created with tools like Sora AI. Understanding and addressing these effects is essential to ensure that social media remains a trustworthy and reliable space for communication and information sharing.

C. Expected Outcomes

Through this project, we expect to produce both quantitative and visual results that clearly show how misinformation flows through X (Twitter). We will create network graphs that map users and interactions to reveal misinformation communities and their connections, along with influence rankings that identify key "source" and "hub" accounts using HITS and PageRank. Cascade dynamics analyses will show how quickly and how far misinformation spreads compared to factual news, while comparative metrics will highlight structural differences such as clustering and modularity, between fake and real news networks. These findings will be supported by visualizations and insights, including graphs that illustrate how misinformation clusters evolve over time or which nodes trigger large cascades. Overall, the project aims to reveal who drives misinformation, how it clusters, and how its spread differs from factual content. Ultimately, we hope this analysis will encourage users to fact-check information before sharing it and provide valuable insights for improving moderation and detection systems.

II. Formalization into a Data Mining Task

A. Data Representation

The primary data type is a directed, time-stamped social graph where nodes represent users and edges represent interactions (retweets, replies, mentions) annotated with timestamps. Node and edge attributes include user metadata, interaction counts, and tweet/text representations such as TF-IDF or BERT embeddings; cascades can optionally be labeled for membership and as fake/real sources. Supplemental data objects consist of a tweet embedding matrix (for example, reduced with SVD or indexed with LSH) and per-cascade timelines used for temporal mining and diffusion analysis.

B. Methodologies

The project will use several analytical methods to model misinformation spread. Link analysis (PageRank, HITS) will identify key sources and amplifiers. Community detection (Louvain or Girvan–Newman) will reveal clusters of misinformation within the user graph. Temporal modeling using the Independent Cascade Model (ICM) and cascade metrics (depth, width, spread rate) will track diffusion over time. If time permits, dimensionality reduction with Truncated SVD and LSH will be applied to speed up clustering and similarity searches in high-dimensional text data.

III. Data Plan

A. Data Description

We will use social network and news data that include both textual and interaction information. Each data sample will consist of the news content (headline, body text, and publisher), user interactions (retweets, replies, mentions, and timestamps), and user profile metadata (followers, followees, and activity level). Additionally, each article or tweet will be labeled as either fake or real news. This combination enables the construction of a directed, time-stamped social graph that captures both the content and the diffusion behavior of misinformation.

B. Data Sources

We will obtain our data from two main sources: the Twitter API and the FakeNewsNet dataset, an open-source repository developed by Arizona State University for fake-news research. FakeNewsNet provides publicly available data in JSON format, which includes both social media interactions and news content. We will download the dataset directly from the repository and, if needed, supplement it with additional data collected through the Twitter API to capture recent or specific misinformation trends.

IV. Implementation

The project will be implemented in several stages, combining data collection, preprocessing, graph construction, and network analysis using established data mining algorithms. Our primary focus is to model the spread of misinformation through user interactions on X (Twitter) using the FakeNewsNet dataset.

A. Data Acquisition and Preparation

We will begin by downloading the FakeNewsNet dataset from its official GitHub repository. This dataset contains both news content (headline, body text, publisher, and labels for fake or real news) and social context (user profiles, followers, followees, posts, and interactions). The social context will allow us to construct a directed, time-stamped social graph, where each node represents a user and each edge represents an interaction such as a retweet, reply, or mention.

During preprocessing, all tweet text will be cleaned by removing links, special characters, and stop words. We will extract timestamps for temporal analysis and apply TF-IDF or BERT embeddings to represent textual content numerically. This step ensures that both the content and structure of misinformation cascades are captured for analysis.

B. Graph Construction

After data cleaning, we will build a global interaction graph $G = (V, E)$ to represent how information flows among users.

Nodes (V): represent users who post, reply, or retweet.

Edges (E): directed connections representing the direction of information flow, annotated with timestamps.

Each node will be associated with metadata such as user activity level, follower count, and average retweet frequency. The edges will also be labeled to indicate whether the interaction is linked to fake or real news, allowing us to compare diffusion behaviors between the two.

C. Link Analysis for Influence Detection

We will use HITS and PageRank algorithms to identify the most influential users in the misinformation network.

HITS will assign two scores to each user:

- Authority score (how often a user's posts are shared), and
- Hub score (how often a user shares others' posts).

PageRank will measure the overall influence of each node within the network, considering both incoming and outgoing connections.

This step will reveal who starts and who amplifies misinformation, highlighting top sources and hubs.

D. Community Detection

To identify groups of users who frequently interact and share similar misinformation, we will apply Louvain community detection. This algorithm partitions the graph into communities that maximize modularity, meaning users inside a group are more connected to each other than to users outside it. We will compare the structure and density of these communities between fake and real news subgraphs to identify echo chambers or coordinated groups spreading misinformation.

E. Temporal Cascade Modeling

We will then model how misinformation spreads over time using the Independent Cascade Model (ICM). In this model, once a user shares fake news, their connected followers have a probability of doing the same. By simulating and analyzing these cascades, we can measure:

- Cascade depth: How many "levels" the misinformation reached,
- Cascade width: How many users were influenced, and
- Spread rate: How quickly misinformation propagated.

Comparing these metrics for fake versus real news will help us understand why false information often spreads more virally.

F. Evaluation and Visualization

Finally, we will evaluate our findings both quantitatively and visually. Metrics such as authority and hub scores, modularity, and cascade statistics (depth, width, speed) will be compared between fake and real news networks. Visualizations such as network graphs, cascade timelines, and community maps will be generated to illustrate how misinformation clusters form and evolve over time.

V. Schedule

Table I presents our project timeline spanning from Week 8 through Week 15, detailing the specific tasks and objectives for each week.

TABLE I: Schedule and Tasks

Week	**Tasks/Objective**
Week 8	Finalize team and submit proposal to Gradescope. Set up repo, environment, and roles.
Week 9	Download FakeNewsNet and start with Twitter API data collection. We'll collect recent tweets using X (Twitter's) API. Retrieving fact-checked labels from FactCheck.org, Snopes.com (The definitive fact-checking site and reference source for urban legends, folklore, myths, rumors, and misinformation), Fake News Detection Datasets
Week 10	Data Preparation & Graph Build: Parsing data-News content and social context. Cleaning text, creating tweet embeddings, building global directed graph and building per-article cascade subgraphs.
Week 11	Link Analysis & First Visuals: Running PageRank and HITS (authority/hub) on global graph, fake subgraph, real subgraph. Producing top-k influencer tables and initial network plots.
Week 12	Community Detection: Running Louvain (or Girvan–Newman if needed). Computing modularity, community sizes, density. Comparing fake-dominant vs real-dominant communities. Visualizing community maps. Dimensionality Reduction (SVD / LSH Task).
Week 13	Evaluation & Comparative Analysis: PageRank/HITS distributions, modularity, assortativity, statistical tests, finalizing figures.
Week 14	Write a formal report: Introduction, Related work, Data, Methods, Results, and Discussion.
Week 15	Building presentation, Short demo. Submit final report, code, and dataset.

VI. Conclusion

This project aims to provide comprehensive insights into the mechanisms of misinformation spread on X (Twitter) through rigorous network analysis and data mining techniques. By identifying key influencers, communities, and temporal patterns, we hope to contribute to the development of better detection and moderation systems while encouraging users to practice critical evaluation of information before sharing.