	DEPARTMENT OF COMPUTER ENGINEERING

Experiment No. 04

Semester	B.E. Semester VIII – Computer Engineering
Subject	Social Media Analytics
Subject Professor In-charge	Prof. Amit Alyani
Academic Year	2024-25

Student Name	Deep Salunkhe 21102A0014
Student Name	Sahil Pokharkar 21102A0006
Student Name	Omkar Patil 21102A0003
Student Name	Pranav Redij 21102A0005

Topic

Analyzing and Visualizing Twitter Social Graph: A Network-Based Approach

---

Domain/Area

- Social Media Analytics

- Graph Theory & Visualization

- Natural Language Processing (NLP)

- Network Analysis

Paper Details Studied

1. **"Understanding Information Diffusion in Social Networks"** - Analyzing how information spreads through connections.

2. **"Graph-Based Approaches for Social Media Analysis"** - Techniques for creating and analyzing graphs from social media data.

3. **"Sentiment and Influence Analysis in Twitter Networks"** - Understanding user influence based on engagement and sentiment.

4. **"NetworkX: A Tool for Network Analysis"** - A guide to using NetworkX for social graph analysis.

Data Set Information

Data Source:

- **Web Scraping using JavaScript in the Browser Console**: Manually extracting tweets, users, mentions, retweets, and replies.

Data Attributes:

- **Users (Nodes)**: Twitter handles, usernames, follower count.

- **Tweets (Nodes)**: Tweet ID, text, timestamp.

- **Edges (Connections):**

  o **Mentions (@username)** → Directed edges from author to mentioned users.

  o **Retweets** → Directed edges from retweeter to original tweet.

  o **Replies** → Directed edges from replier to the original poster.

Flow of Project

Step 1: Data Collection

- Use **JavaScript in the browser console** to scrape Twitter data.

- Extract tweets, user details, mentions, replies, retweets.

- Store data in **JSON or CSV format**.

Step 2: Preprocessing

- Remove duplicate or irrelevant tweets.

- Structure data into **Nodes (users, tweets)** and **Edges (interactions)**.

Step 3: Graph Construction

- Use **NetworkX** to create a directed graph.

- **Nodes**: Users & Tweets.

- **Edges**: Mentions, Replies, Retweets.

- Assign **weights** to edges based on interaction frequency.

Step 4: Graph Visualization

- Use **Matplotlib & NetworkX** to plot the graph.

- Implement **Gephi or Plotly** for interactive visualization.

- Differentiate nodes by:

  o Color: User influence (high-degree nodes = darker color).

  o Size: Number of interactions.

Step 5: Analysis & Insights

- **Find Key Influencers:** Degree Centrality, Betweenness Centrality.

- **Community Detection:** Apply **Louvain Algorithm** to identify communities.

- **Virality Analysis:** Identify tweets with the most retweets and mentions.

- **Sentiment Analysis (Optional):** Use **NLTK or VADER** to analyze tweet sentiment.

Step 6: Conclusion & Future Scope

- Summarize key findings.

- Discuss possible improvements, such as:

  o Expanding to multiple social platforms (Instagram, Reddit).

  o Adding NLP-based topic modeling.

o Using **Machine Learning** for bot detection.

Tools & Technologies Used

- **Data Collection**: JavaScript (for scraping)

- **Graph Processing**: NetworkX, Pandas

- **Visualization**: Matplotlib, Plotly, Gephi

- **Storage**: JSON, CSV, PostgreSQL (Optional)

- **Analysis**: NLP (NLTK, VADER), Scikit-learn (for clustering)

Expected Outcome

- A **graph-based visualization** of Twitter interactions.

- Identification of **key influencers** in a topic-based network.

- Detection of **social communities** based on interaction patterns.

- Insights into how information spreads through Twitter networks.

Conclusion

This project aims to provide **meaningful insights into social interactions on Twitter** by visualizing user connections and analyzing engagement patterns. The application of graph theory and analytics will help uncover **influential users, trending topics, and network communities**, making it useful for social media marketers, researchers, and analysts.