Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Subject: Social Network Analysis Laboratory (DJ19DSL8014)

AY: 2024-25

Bhuvi Ghosh
60009210191

Experiment 9

Aim: To study information diffusion models through real-time case analysis, focusing on the dynamics of how information spreads in a social network.

Theory:

Information diffusion models explain the process through which information propagates in networks such as social media, communication systems, and biological environments. These models help understand the mechanisms of influence, adoption, and spread, aiding in fields like marketing, epidemiology, and cybersecurity.

Core Models:

Independent Cascade Model: Each node influenced has a chance to spread the information to its neighbors based on a probability factor.

Linear Threshold Model: A node adopts the information if the cumulative influence from its neighbors exceeds a certain threshold.

SIR Model (Susceptible-Infectious-Recovered): Commonly used in epidemiology, this model tracks nodes from being susceptible to becoming infectious and eventually recovering.

Metrics for Analysis:

Spread Size: The number of nodes that received the information.

Time of Spread: How quickly information reaches the majority.

Influence Maximization: Identifying the most influential nodes to maximize spread.

Case Study: Tracking Viral Tweets on Social Media Platforms

Objective:

To understand how tweets become viral on social media using information diffusion models. The study focuses on identifying the key factors and users responsible for a tweet's widespread dissemination and analyzing the pattern and speed of its spread.

Step-by-Step Breakdown:

1. Data Collection:

The first step was to gather data related to a set of tweets that went viral. This involved:

- Extracting metadata from the Twitter API such as:
    - Tweet content
    - Number of retweets, likes, and replies
    - Timestamps
    - User information (followers, verification status, etc.)
    - Mentions, hashtags, and URLs
- Building a **social graph**, where:
    - Nodes represent users
    - Edges represent interactions (e.g., retweets, mentions, replies)

This graph served as the basis for modeling the spread of information.

---

2. Applying Diffusion Models:

Two primary models were applied:

A. Independent Cascade (IC) Model:

- Assumes that once a user tweets or retweets, they have **one chance** to influence each of their followers (neighbors).
- Each edge has a **probability** (e.g., 0.1 or 10%) that the information will spread through it.
- The process continues **step-by-step** until no further activations occur.

B. Linear Threshold (LT) Model:

- Each user has a **threshold** value representing how easily they are influenced.
- When the **combined influence** from their active (retweeting) neighbors **exceeds this threshold**, they retweet the tweet.
- This model is more deterministic and reflects peer pressure or cumulative influence.

---

3. Analysis Using Metrics:

The simulation using both models yielded the following key metrics:

a. Spread Size:

- Counted how many users ended up seeing or retweeting the original tweet.
- Larger spread size indicates higher virality.

b. Time of Spread:

- Measured how quickly the tweet was shared over time.

- Plotted spread over time to observe if it peaked instantly (breaking news) or grew gradually.

c. Influence Maximization:

- Identified **users who, if influenced initially**, caused the largest spread.
- These were often:
  - Verified accounts
  - Users with a high follower count
  - Accounts known for timely or sensational content

Used centrality metrics (like degree, closeness, and betweenness) to validate influencer roles.

Key Observations and Findings:

1. Key Influencers Identified:

- Verified accounts and celebrities had a **disproportionately large influence** in spreading tweets.
- When such accounts retweeted a post, it rapidly gained visibility and engagement.

2. Time-Sensitive Spreading:

- Tweets concerning **breaking news, trending topics, or emergencies** spread significantly faster.
- The **initial few minutes** played a crucial role in determining virality.
- After a few hours, the rate of spread typically dropped as newer tweets replaced them.

3. Cluster Effects:

- In **tight-knit communities** (like fandoms or regional groups), tweets spread quickly within the group.
- These clusters helped amplify tweets before they crossed over to a larger network.
- Community detection algorithms showed that **dense clusters** contributed to early-stage virality.

Real-Life Implications:

- **Marketing:** Brands can use such analysis to **launch products or campaigns** by targeting influencers in specific clusters.
- **Misinformation Control:** Early identification of fake news spread can help in **limiting damage**.
- **Content Strategy:** Platforms can **recommend content** to high-influence users to maximize reach.

Conclusion:

Information diffusion models provided valuable insights into social media dynamics, allowing for improved content targeting, marketing strategies, and misinformation control.