# EXPLORATORY PROJECT

# RUMOR MINIMIZATION IN SOCIAL NETWORKS

**Name-** Palakur Eshwitha Sai

**Roll No-** 18123014

**Branch-** Mathematics and Computing

**Supervised By-** Dr. Lavanya Selvaganesh

# ABSTRACT

- Online social networks have become one of the most important medium to propagate the information among communities. People are able to share any kind of knowledge with each other online.
- In today's information-oriented society, a mechanism to suppress harmful rumors in social web has become very important.
- In this project, we study propose static and dynamic algorithms for rumor minimization.
- Our goal is to minimize the influence of the rumor (i.e., the number of users that have accepted and sent the rumor) by blocking a certain subset of nodes.[6]


## INTRODUCTION

- If any information circulates without officially publicized confirmation, it is called a rumor.
- Using social networks, people can share their innovations, ideas, happening topics, etc. On the other hand, they may become the channel for spreading rumors too.
- Rumor blocking is a serious problem in large-scale social networks. Malicious rumors could cause chaos in society and hence need to be blocked as soon as possible after being detected.
- Before going into the details of how this can be achieved, let us know more about social networks and their analysis using graph theory and network analysis.
- For further details we refer to [3],[4,],[5],[6]

Let us learn know about few essential terms related to graph theory and networks before learning more about social network analysis.

## TERMINOLOGY

- **Graphs and Networks**

– **What is a graph?**

  - A **Graph** is a non-linear data structure consisting of a finite set of vertices (or **nodes**) and set of **edges** which connect a pair of nodes. They are used to study pairwise relationship between entities.
  - **Vertex**

    Set of objects (also called nodes) that are connected together.

  - **Edge**

    The connection between the nodes are called edges or links.

  - **Degree**

    Degree of a node represents the number of links it has to other nodes. The degree can represent the number of different individuals a person has talked to in his call graph.

– **Representations of Graph**

  - **Adjacency matrix** and **Adjacency list** are used to represent a graph showing it' s nodes and links between them.
  - **Adjacency matrix**

    Adjacency Matrix is a 2D array of size V x V where V is the number of vertices in a graph. Let the 2D array be adj[][], a slot adj[i][j] = 1 indicates that there is an edge from vertex i to vertex j. Adjacency Matrix is also used to represent weighted graphs. If adj[i][j] = w, then there is an edge from vertex i to vertex j with weight w.

- **Adjacency list**

  An array of lists is used. Size of the array is equal to the number of vertices. Let the array be array[]. An entry array[i] represents the list of vertices adjacent to the ith vertex. This representation can also be used to represent a weighted graph. The weights of edges can be represented as lists of pairs.[1]

– **What is a network?**
  - A **network** is a catalog of a system's components often called **nodes** or **vertices** and the direct interactions between them, called **links** or **edges.**
  - If the edges in a network are directed, i.e., pointing only in one direction, it is called a **directed network.**
  - If all the edges are bidirectional or undirected, then it is called an **undirected network.**

– **Applications of graphs**
  - Graphs are used to solve many real-life problems. They provide a better way of dealing with abstract concepts like relationships and interactions.
  - Graphs are used to represent networks. The networks may include paths in a city or telephone network or circuit network.
  - Graphs are also used in social networks like Facebook. For example, in Facebook, each person is represented with a vertex (or node). Each node is a structure and contains information like person id, name, gender, etc.

## What is a social network?

- A social network, in mathematical context, can be formulated as a directed graph G = (V, E) consisting of a set of nodes V representing the users, and a set of directed edges E denoting the relationship between users (e.g. following or being followed).
- Let |V | = N denote the number of nodes, and (u, v) ∈ E denote the directed edge from node u to node v (u, v ∈ V), and $\alpha_{uv} \in \{0, 1\}$ denote the edge coefficient, where $\alpha_{uv} = 1$ represents the existence of edge (u, v), and $\alpha_{uv} = 0$, otherwise.
- We use $p_{uv}$ to denote the probability of u sending the rumor to v and v accepting it, i.e., the success probability of u activating v.
- Let D(u) denote the in-degree of node u that is number of edges coming towards node u.[3]

## SOCIAL NETWORK ANALYSIS

- Social Network Analysis (SNA) is the process of investigating social structures through the use of networks and graph theory.
- It characterizes networked structures in terms of nodes (individual people or things within the network) and the edges (interactions) that connect them.
- These networks are often visualized through socio-grams in which nodes are represented as points and edges are represented as lines which help us to assess the networks qualitatively by varying the visual representation of their nodes and edges to reflect attributes of interest.
- Social network researchers measure network activity for a node by using the concept of degrees -- the number of direct connections a node has (Degree Centrality).[7]

- Social network analysis [**SNA**] is the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information/knowledge entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes.

## INFORMATION DIFFUSION MODELS

- Diffusion models describe the process of information propagating through the network.
- Two classic diffusion models are Linear Threshold (LT) and Independent Cascade (IC) model.

- **LT MODEL**
    - In LT model, each node in the network selects a random threshold value, then at each time step inactive node gets influenced by all its active neighbors.
    - The inactive node turns into active once the total weight of all its incoming neighbors reaches a threshold value.
    - Similar to IC model, LT model is majorly utilized to maximize the influence of spread in network.

- **IC MODEL**
    - The whole propagation process proceeds in discrete time steps $t_0$, $t_1$, $t_2$ and so on.
    - Initially, the cascade is triggered by a set of activated nodes, i.e., the *seed nodes* at $t_0$.
    - At one given instance, the node has only one chance to activate one of its inactive neighboring nodes.

    - The IC model describes the flow of information from source to other nodes in terms of a directed graph, which is helpful for finding highly

influential users and identifying a source of misinformation by analyzing the diffusion network in reverse direction.
  - IC model is also used for finding the maximum influence.[8]

# RUMOUR DIFFUSION MODELS

## 1. MACROSCOPIC OR EPIDEMIC MODELS

- Rumor diffusion mechanism is similar with that of epidemic propagation.
- During the propagation of rumors, each node could have one of the following three states: Susceptible (S), Infected (I) and Recovered (R), which is known as the SIR model.
- **Susceptible-** The state of being susceptible represents the node has the potential to accept and spread the rumor at any time (also called Spreader).
- **Infected-** Infected indicates the node has already accepted and spread the rumor (also called Ignorant).
- **Recovered-** Recovered denotes the state of the node identifying the rumor and denying it (also called Stifler ). In a social network, the recovered node is the one who is aware of a rumor, therefore, either it will delete the post or will not forward that post to a neighboring node.
- The rumor is propagated through the population by pair-wise contacts between spreaders and others in the population.
- One famous variant is Maki-Thompson (MK) model. Three types of interactions can happen with certain rates-
  - When a spreader meets an ignorant, the ignorant will become a spreader.
  - When two spreaders meet with each other, one of them will become a stifler.
  - When a spreader meets a stifler, the spreader will lose the interest in spreading the rumor and hence becomes a stifler.

- The other basic epidemic models are SI, SIS and SIRS.
- **SI model-** Initially nodes are susceptible (S) and can be infected (I) with propagated rumors in the SI model. Susceptible nodes are uninfected nodes having infected neighbors and therefore have a higher probability of becoming infected and a contaminated node remains infected forever. SI model is not viable as it does not take into account that contaminated users can be cured after having been contaminated.
- **SIS model-** In the SIS model, the probable states are yet again susceptible (S) and infected (I), but in this model, when susceptible nodes become contaminated they can cure back to susceptible after some period.
- **SIRS model-** In SIRS model, a recovered node can again become a susceptible node with some probability.[8]

## 2. MICROSCOPIC MODELS

- The known models in this category are Independent Cascades (IC) and Linear Threshold (LT) Models which are explained above.
- The microscopic approaches attracted more attention in the individual's interaction: “who influenced whom.”
- Both these models are majorly used in influence maximisation.

Most of the work concentrates on maximizing the positive information through social networks based on the IC (Independent Cascade) model. On the other side, the negative influence or rumor minimization problem has less attention. Now, let us look into some solutions for the latter.[8]

## RUMOR INFLUENCE MINIMIZATION

- Rumor influence minimization addresses the problem of minimizing the propagation effect of undesirable rumors in social networks.
- The rumor influence minimization problem is converse to the classic influence maximization problem.
- Different rumor minimization problems are defined based on different definitions of the network.
- Here, we try to study two different algorithms for rumor minimization in the SIR model of rumor diffusion proposed above.

## ALGORITHMS FOR RUMOR MINIMIZATION

The following rumor minimization algorithms are developed for the SIR model-

1. **GREEDY ALGORITHM**
    - Let $A_0$ be the original network coefficient matrix before any nodes are blocked.
    - The proposed Greedy algorithm tries to block the rumor as fast as possible to prevent the rumor from further propagation.
    - The working mechanism is as follows- At time $t_0$ when we detect the rumor, we immediately select all K nodes in our budget and block them (i.e., remove all the links of it so that it cannot communicate with its neighbors).
    - Mathematically, the Greedy algorithm aims to minimize the likelihood of inactive nodes getting activated at t1, i.e., the next time stamp after the rumor is detected.
    - It is a static blocking algorithm.[6]

## 2. DYNAMIC BLOCKING ALGORITHM

- Greedy algorithm is a static type of blocking algorithm.
- Different from the greedy blocking algorithm, the dynamic blocking algorithm blocks the K nodes in separated steps.
- In this case, the blocking strategy is split into several rounds and each round can be regarded as a greedy algorithm.
- Instead of blocking K candidates at the moment of detection, as in the previous static blocking strategy, this dynamic approach is carried out in a progressive way.
- The working of the dynamic algorithm is as follows-at the very first stage of blocking, we select a number $k_1$ nodes to block based on the Edge matrix and previously infected nodes.
- In the next round, we move forward with the rumor diffusion, and then use the updated status to block additional $k_2$ nodes. The blocking process continues at each following instants until the budget runs out at a moment t, which can be expressed as $\Sigma k_j = K$.
- In real implementation, we decrease k as time goes by, and a practical example is $k_j$=2^(-j)*K.
- The activation likelihood of a given moment is a variable which depends on the temporal Edge matrix and previous status. So, we take advantage of instantaneous information all through the process,
- Rather than sparing all the efforts at once, we apply consequent force to block the diffusion of rumors. So, this method is more efficient than the static methods.[6]

**CONCLUSION-**

In this project, we focused on the rumor minimization problem in social networks and studied two algorithms to achieve the same. We observe that, the dynamic algorithm is more efficient than the static greedy algorithm. Both of these algorithms disconnect the affected nodes so that the rumor spread is minimized.

**REFERENCES-**


1. https://www.geeksforgeeks.org/graph-and-its-representations/
2. https://phys.org/news/2019-06-explores-gossip-social-networks.html
3. Ms. R. Amutha and Dr. D. Vimal Kumar, Detection and Minimization of Rumor Influence in Social Networks, International Journal of Computer Sciences and Engineering, Vol.-7, Issue-3, March-2019, Page No: 942-945.
4. Surekha Rawale, Sapana Birkure, Kajal Hiray , Prof. Deepti Deshmukh, Detection and Minimization Influence of Rumor in Social Network, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 11, Nov –2017, Page No. 757-761.
5. ANURAG SINGH, Rumor Dynamics and Inoculation of Nodes in Complex Networks, Thesis, Department of Electrical Engineering, IIT Kanpur.
6. Iao Wang, Ge Chen, Luoyi Fu, Li Song, Xinbing Wang, Xue Liu, DRIMUX: Dynamic Rumor Influence Minimization with User Experience in Social Networks, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Page No. 791-797.
7. https://en.wikipedia.org/wiki/Social_network_analysis
8. Sushila Shelke and Vahida Attar, Source detection of rumor in social network – A review, Online Social Networks and Media, Volume. 9, January 2019, Page No.30-42