A Web Based Application for Simulating and Analyzing the Dynamics of Conflicting Information Spread in Social Networks

1st Soumil Sinha
Department of Computer Science and Engineering
Indian Institute of Technology (IIT) Jodhpur
Jodhpur, India
sinha.5@iitj.ac.in

2nd Jahnab Dutta
Department of Computer Science and Engineering
Indian Institute of Technology (IIT) Jodhpur
Jodhpur, India
dutta.4@iitj.ac.in

3rd Anggy Eka Pratiwi
Department of Computer Science and Engineering
Indian Institute of Technology (IIT) Jodhpur
Jodhpur, India
pratiwi.1@iitj.ac.in

4th Suman Kundu
Department of Computer Science and Engineering
Indian Institute of Technology (IIT) Jodhpur
Jodhpur, India
suman@iitj.ac.in

***Abstract*—Social networks have become vital communication platforms, enabling rapid global sharing of news, ideas, and cultural content. Their digital framework facilitates information spread and enhances user interaction, fostering communities and encouraging content to go viral quickly. Within this context, the theory of information diffusion becomes particularly relevant. It explains how information spreads through the social interaction of these networks, which is influenced by the strength of connections between users and the structural characteristics of the network itself. This paper analyzes information diffusion, mainly focusing on competitive information within such networks. Specifically, it explores scenarios in a network that would adopt only one of two types of information: type A and type B. Various factors play a role in determining how this plays out: people in the network might be biased towards one type of information or another, the level of trust between a person and the people it is connected to, the decaying freshness of information in the network, and more. By adopting the Susceptible-Infected-Recovered (SIR) model, this research aims to provide a complex understanding of how information propagates in differing social network contexts. We develop a web-based interface that enhances the research framework by enabling asynchronous submission of simulation tasks and allowing precise adjustments of the simulation parameters to optimize the effectiveness of the models. The interface facilitates extensive experimentation across varied network structures, offering insights into strategic information dissemination practices.**

***Index Terms*—Competitive information diffusion, information diffusion, and social networks.**

I. Introduction

In recent years, there has been a significant growth in the popularity of social networking apps like Facebook, Instagram, and Whats-App [1]. These platforms have become crucial for spreading and receiving various types of information [2] [3] [4]. Therefore, it is important to develop models using graphs to analyze how information spreads and diffuses dynamically. Information diffusion models are one of the analytical tools that can help researchers understand the patterns and dynamics of how information spreads across various networks. Information diffusion models provide foundational frameworks for understanding how information, ideas, or behaviors spread through social systems. These models are widely used in various fields, including sociology, marketing, public health, and computer science. The SIR model is one of the most influential and classic models [7].

SIR MODEL

Kermack [5] proposed the SIR model to look into the dynamics of epidemics. Typically, there are three possible states for each individual: susceptible (S), infected (I) and removed (R). During each interaction, the individual with a susceptible state will switch to an infected state with probability α if he interacts with infected individuals, while an individual with a research state can become recovered and switch to an R state with probability γ. It is assumed that the individual in removed state will keep this state unchanged forever. Researchers frequently use SIR-based models to study the dynamics of information diffusion, and these models assume that the three possible states are ignorance, spreader, and stiflers , respectively. Most of the SIR-based research on information diffusion considers a single piece of information. However, the case when two competitive pieces of information (denoted by A and B) are spreading at the same time has been studied by Fu [8].

MODIFIED SIR

Fu [8] introduced four possible states in the Modified-SIR (MSIR) for each individual: ignorant (I), spreader A (S_A),

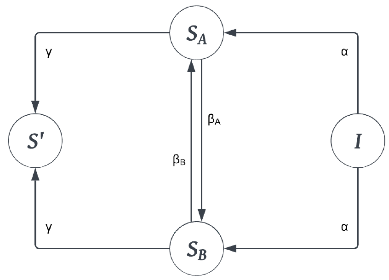

Fig. 1. State Diagram for the M-SIR Model [8]

spreader B (S_B), and stiflers (S') are nodes that removed from information spread [9]. The state diagram for the individual in the MSIR is shown in Figure 1. The rule for an individual spreading information over social networks according to this model can be described as follows: (i) Originally, there were some seeds in the population carrying information A and B, respectively, while other individuals are ignorant. (ii) The ignorant individual turns into a spreader A (B) with probability α when interacts with spreader A(B); otherwise, will keep the ignorant state unchanged. (iii) The spreader A(B) may change his state to spreader B(A) with probability β_A (β_B) when interacts with spread B(A). (iv) The spreader may lose interest in both information A and B and eventually become stiflers with probability γ . Although this model only considers two pieces of information spreading over social networks, it can be easily extended to the case of more than two pieces of information.

Contribution: Our work significantly extends the research outlined in Fu [8]. This research begins by defining the problem statement, which consists of modeling the social network using a graph, followed by the properties of the nodes that try to model human behavior, and finally discusses the halting conditions of information spread. The methodology will provide a detailed discussion of the exact algorithm of the simulation as well as the implementation of the simulation. The results will be presented for various large social networks that are well-known in the literature.

II. Problem

A. Network Structure

For our social network, we will consider a graph consisting of N nodes, each representing an individual in the network. These nodes are denoted by $V = 0, 1, \ldots, N-1$, and the set of edges in this network is denoted by $E \subset V \times V$. Each of these edges is a link between two nodes, and we can also define the set of neighbors for a node $i \in V$ as $N_i = \{j \in V \mid (i, j) \in E\}$. We simulate the diffusion of competitive information in this social network by randomly choosing some spreaders of information A and B. Other individuals are exposed to two types of information and have the choice to adopt (and consequently spread) it or ignore it. We will assume that the spreading of information follows the SIR model.

B. Node Properties

There are three possible states for an individual in the SIR Model: ignorant (I), spreader (S), and stifler (S'). Additionally, we divide the population into three subgroups: innovators, ordinary and conservatives. This will directly affect α, since an innovator would be more likely to spread information than an ordinary individual and even more likely than a conservative individual. This is determined with the help of an additional parameter $\theta \in (0, 1)$, which is randomly chosen for each individual. Then, $\alpha(\theta)$ is calculated as:

$$\text{For innovative individuals:} \quad \alpha = \min(1, \alpha_0(1+\theta)) \tag{1}$$

$$\text{For ordinary individuals:} \quad \alpha = \alpha_0 \tag{2}$$

$$\text{For conservative individuals:} \quad \alpha = \alpha_0\theta \tag{3}$$

where α_0 is the initial spreading rate assigned to all nodes. Similarly, θ also affects β directly since an innovator is more likely to switch the type of information it is spreading than an ordinary person and even more likely than a conservative person. Hence, $\beta(\theta)$ is calculated as:

$$\text{For innovative individuals:} \quad \beta = \min(1, \beta_0(1+\theta)) \tag{4}$$

$$\text{For ordinary individuals:} \quad \beta = \beta_0 \tag{5}$$

$$\text{For conservative individuals:} \quad \beta = \beta_0\theta \tag{6}$$

where β_0 is the initial switching rate assigned to all nodes. The above formula for β_A and β_B. Apart from these spreading and switching rates, we have also considered the bias of an individual towards information A or B. For example, an Indian person is browsing some form of social media when he sees a post informing him that India will win the match against Pakistan tomorrow. He is most likely biased towards this information and will spread it further, as it favors India. If he receives the information that Pakistan has a chance of winning, then he will probably reject that piece of information because of his own bias. The modeling has been done using $\delta \in [0, 1]$. If this value is 0, the individual is seen as completely biased towards information A; if the value is 1, a complete bias towards information B is exhibited by the individual. We also consider how trust between different nodes affects the change in their states. This can be illustrated in the sense that a person has various levels of trust for each source of information. We are more likely to trust a post by the official board of control for cricket website than a random Indian cricket fan page. To model this, we use $l_{ij} \in (0, 1)$, which represents the level of trust link strength) between nodes i and j. Now that we have these two parameters established, let's see how the probability of a node i changing its state from an ignorant to a spreader when it receives new information from its neighbor j is calculated: If information A is received:

$$P(\text{Ignorant} \rightarrow Spreader_A) = \alpha(\theta) \times l_{ij} \times (1-\delta) \tag{7}$$

Otherwise, If information B received :

$$P(\text{Ignorant} \rightarrow Spreader_B) = \alpha(\theta) \times l_{ij} \times \delta \tag{8}$$

Thus, we can see how both bias and link strength directly affect the probability of spreading information. There is one additional parameter that we have considered that directly affects the switching probability for a node, namely data freshness. To explain this, consider a situation where a node i is in the state of $Spreader_A$ and none of its neighbors are sharing information B with it. They are only sharing information A. So when information B is shared with this node next time, it would be believed that information B is very old (state) while information A is fresh since its neighbors have been spreading information A in the recent past. Hence, the probability of it switching to information B decreases. On the other hand, if the node has been receiving information B and not A, then it would not switch to B with a probability greater than $\beta_A(\theta)$. The freshness of information A is calculated as follows:

$$y_A = max(0, 1 - 0.01 \times (t - t_A)) \tag{9}$$

Where t_A is the previous timestamp when information A was received (may or may not have been adopted). The switching probabilities are calculated as follows:

$$P(Spreader_A \rightarrow Spreader_B)$$
$$= \min\left(\beta_A(\theta), \beta_A(\theta) \times (1 + \tanh(y_B - y_A))\right) \tag{10}$$

$$P(Spreader_B \rightarrow Spreader_A)$$
$$= \min\left(\beta_B(\theta), \beta_B(\theta) \times (1 + \tanh(y_A - y_B))\right) \tag{11}$$

Since $0 \leq y_A, y_B \leq 1$, we know that $-1 \leq (y_A - y_B) \leq 1$

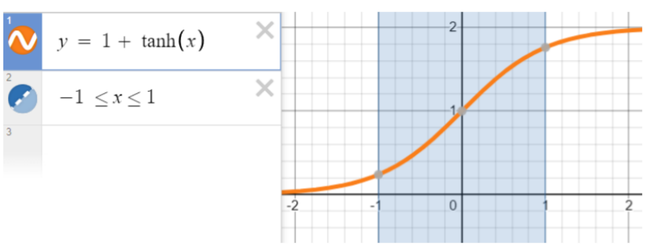

Fig. 2. Function to model switching preference, based on information freshness

We chose this activation function because, in the case of $y_A = y_B$, the switching probabilities remain unaffected, as they should. When deciding whether to switch from A to B, if $y_B < y_A$, then $P(S_A \rightarrow S_B) = \beta_A(\theta)$ note that $P(S_A \rightarrow S_B) \neq 0$ and $P(S_B \rightarrow S_A) \neq 0$

C. Stable Equilibrium

To denote the number of ignorants as $I(t)$, number of spreaders of A as $S_A(t)$, number of spreaders of B as $S_B(t)$ and the number of stifler as $S'(t)$. We know that $I(t) + S_A(t) + S_B(t) + S'(t) = N$, as these are the only possible states for each node. Then,

$$\Delta I = -\alpha \frac{S_A + S_B}{N} I \tag{12}$$

$$\Delta S_A = \alpha \frac{S_A}{N} I - \beta_A \frac{S_B}{N} S_A + \beta_B \frac{S_A}{N} S_B - \gamma S_A \tag{13}$$

$$\Delta S_B = \alpha \frac{S_B}{N} I - \beta_B \frac{S_A}{N} S_B + \beta_A \frac{S_B}{N} S_A - \gamma S_B \tag{14}$$

For an effective stable equilibrium, we set the LHS of these three equations to 0 as we do not want any change in the number of ignorant and spreaders. The condition for this equilibrium comes out to be $SA = SB = 0$, for example information diffusion stops only when there are no more spreaders left in the network. Note that we have considered α, β_A, β_B and γ to be the same for each node in the network, but interestingly, the result also holds for the actual model, where each node will have its own spreading rate α and switching rates β_A and β_B.

III. Approach

A. Evolution of Modified SIR Model

The algorithm outlined in Fu [8] we have implemented and incorporated. The additional parameters discussed in the problem section. Algorithm 1 is the main task. It is being simulated and written inside the simulated function.

Algorithm 1 Information Spread Simulation

```
Initialization of parameters for all nodes
Initialize t = 0
repeat
    for k = 1 to N do
        Choose a random node i ∈ V
        Choose a random neighboring node j which is a spreader
        if no such node exists then
            continue
        end if
        if i is ignorant then
            if j is Spreader_A then
                Update t_A for node i, i.e., latest timestamp for info A
                i becomes Spreader_A with probability P(I → SA)
                Update the number of individuals with different states
            else if j is Spreader_B then
                Update t_B for node i, i.e., latest timestamp for info B
                i becomes Spreader_B with probability P(I → SB)
                Update the number of individuals with different states
            end if
        else if i is a spreader then
            if i and j are spreaders of the same type of information then
                Update t_A (or t_B as the case may be) for node i
            else
                Calculate freshness of info A and B
                if i is Spreader_A then
                    Update t_B for node i, i.e., latest timestamp for info B
                    i becomes Spreader_B with probability P(SA → SB)
                    Update the number of individuals with different states
                else
                    Update t_A for node i, i.e., latest timestamp for info A
                    i becomes Spreader_A with probability P(SB → SA)
                    Update the number of individuals with different states
                end if
            end if
        end if
    end for
    for all spreaders do
        Spreader becomes Stifler with probability γ
        Update the number of individuals with different states
    end for
until number of spreaders becomes 0 or termination time is reached
```

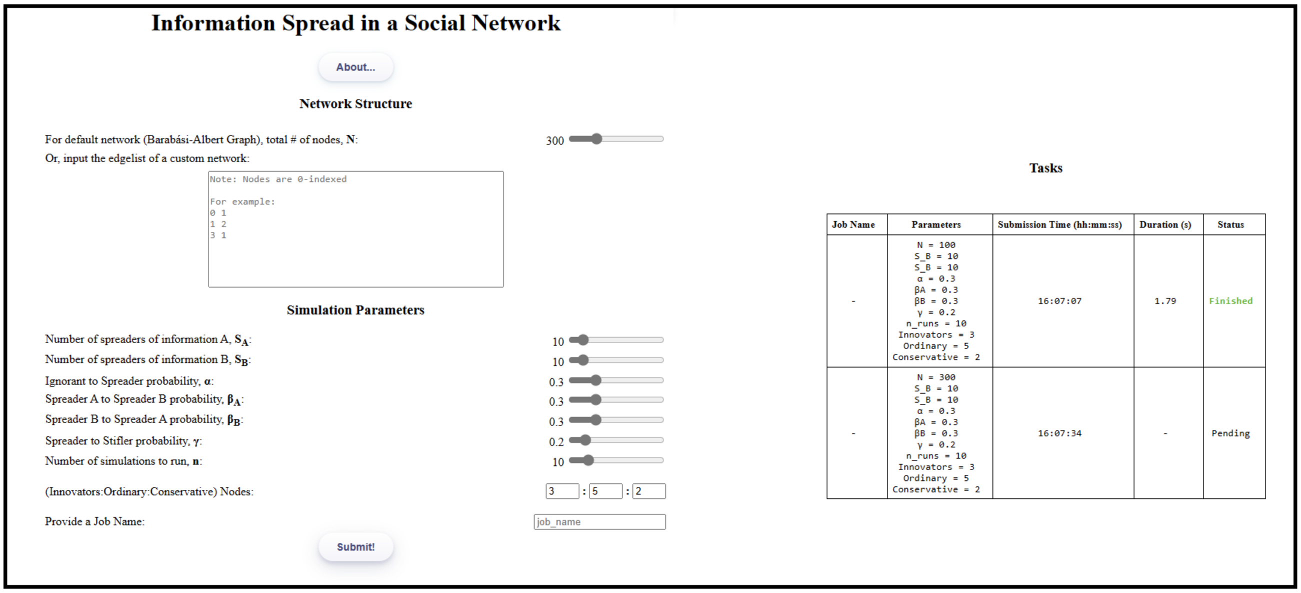

Fig. 3. Simulation application page

B. Simulator Implementation

We have developed a client-server model to create a user-friendly simulator. The simulator logic is implemented in Python, while the server-side logic uses Flask, a library designed to facilitate server-side programming. Flask handles the serving of HTML pages, which allow users to input simulation parameters and to view and analyze the results. Additionally, we use Celery, a library that enables asynchronous task execution in the background via multi threading.

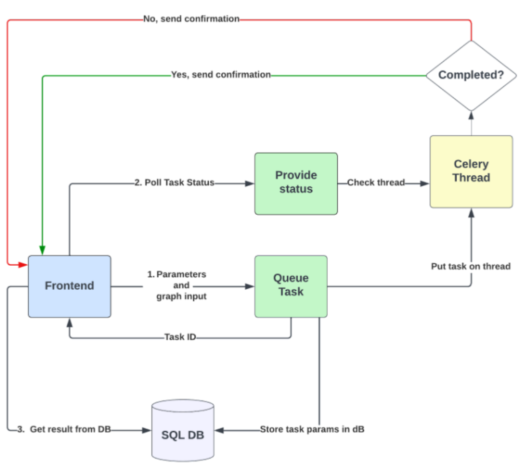

Fig. 4. Simulator Workflow

The workflow is illustrated in Figure 3. The client interacts with the simulator through a screen where parameters and the simulation graph are entered, as shown in Figure 4. After configuring the settings, the client submits these as a POST request to the $'/submitdata'$ URL. The server processes this input by initially storing the parameters and the graph's edge list and using them to create an asynchronous Celery task instead of running the 'simulate' function immediately. The task ID is then returned to the client, and an instance containing this ID, along with submission details, is saved in the SQL database. This instance initially has null-initialized fields for duration and results. In the back end, Celery, using Redis as a broker on port 6379, manages the tasks across configurable threads—default set to four. As the task progresses, its status updates from 'PENDING' to 'SUCCESS' are reflected in the task results table. When the task completes, the server updates the SQL database entry, confirms completion, and the task result table displays the final duration, changing the status to 'Finished'. The user can then access the completed results via a link marked 'Finished' that directs to $'/simresults/ < taskid >'$, where the server retrieves and displays the results on a separate page, with options to download a PDF of the results.

IV. Discussion

We tested our simulator against the standard Barabási-Albert graph with 50, 200, and 1000 nodes. We assume that initially, there are ten spreaders for information A and B. Then, we will test some known large networks, such as the Enron Email Network, the Facebook Combined Social Network, and the Wikipedia "Crocodile" Network. Here, we assume there are 50 spreaders for information A and B. Four plots are drawn to show the change in the proportion of spreaders of each information over time (top-left), the change in the proportion of nodes switching information (A to B or vice versa) over time (top-right), the change in the proportion of nodes changing from spreaders of A/B to stiflers (bottom-left) and finally the change in the proportion of nodes that never received any information from their neighbors over time. All these results are averaged over 50 runs, as we have set n runs to 50.

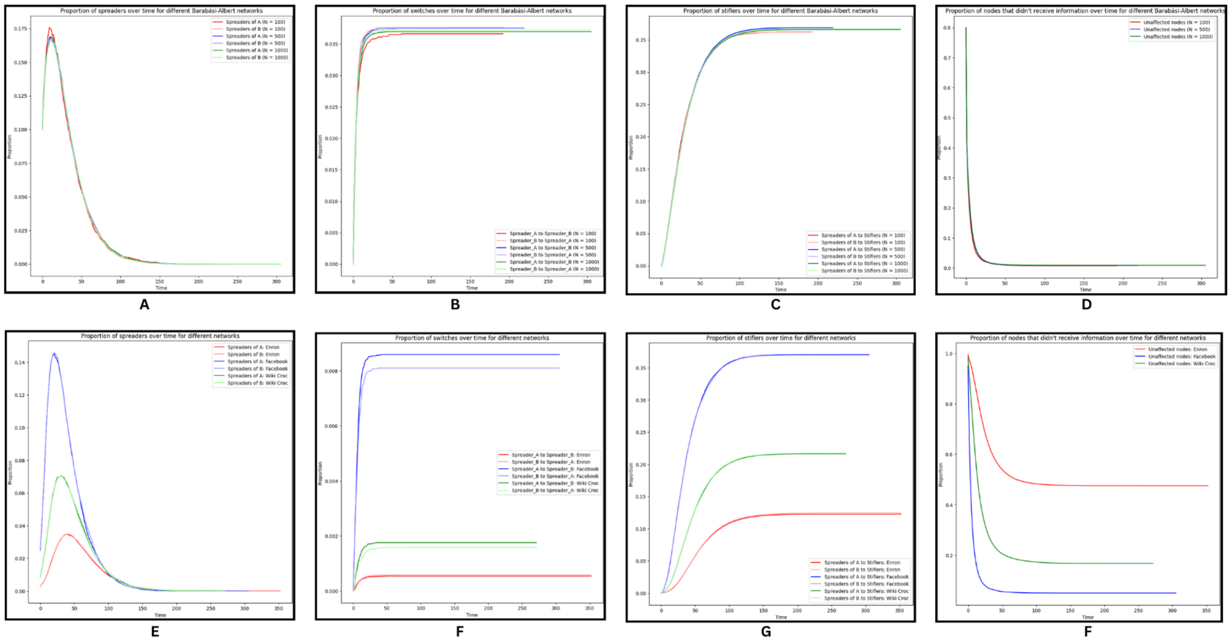

Fig. 5. Simulation Result

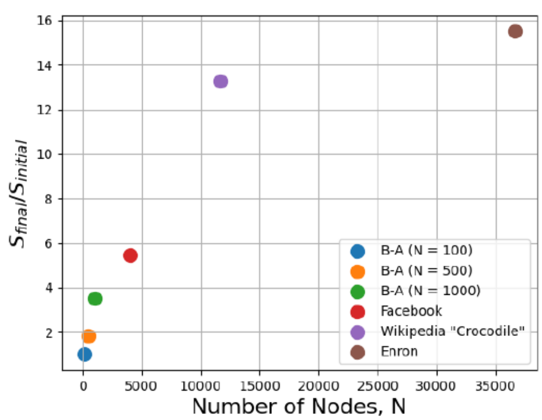

Fig. 6. Simulator Workflow

This simulation provides insights into how different networks facilitate various patterns of information diffusion. For example, sparser networks like the Enron network exhibit a large proportion of individuals who were never exposed to any information, and it also takes the longest time to converge. In contrast, the convergence times for the Facebook and Wikipedia networks are almost identical. Further analysis reveals a correlation between the network size and the maximum number of spreaders: as the number of nodes increases, so does the ratio of the maximum number of spreaders to the initial number of spreaders. This relationship, clearly visible in our plotted data, warrants further investigation. Additionally, despite these differences, all networks generally show similar trends in the dynamics of spreaders, switchers, and stiflers over time, as illustrated in the accompanying plots.

We have analyzed three variants of the Barabási-Albert network: the first with 100 nodes and ten spreaders each for information types A and B; the second with 500 nodes and 50 spreaders for each type; and the third with 1000 nodes and 100 spreaders for each type. We consistently maintained the proportion of initial spreaders across all three networks to ensure comparability. In Figure 5(A), we observe that the proportion of spreaders for each type of information in all three networks follows a similar trajectory of growth and decline. The network with 100 nodes exhibits the highest peak at 0.175, suggesting that smaller networks may facilitate more effective information spread. This implies that if the number of spreaders increases proportionally with the number of nodes, the overall trend of information spread remains consistent, albeit with a slight reduction in the peak spreader proportion as network size increases. Figure 5(B) shows a similar pattern, where the smallest network (100 nodes) has the highest activity. Each network displays a dominant information type, evident from the fluctuating switches between A and B. Notably, the proportion of switchers is highest in the 100-node network, decreasing in the 500-node network, and further in the 1000-node network. Figure 5(C) provides less definitive information but indicates that the 100-node network converges fastest, converting all spreaders to stiflers (the simulation terminates when no spreaders remain). As network size increases, the average convergence time rises, though not directly proportional to the number of nodes. Figure 5(D) reveals that the proportion of unaffected nodes approaches zero in all networks, suggesting that the total number of nodes does not significantly impact the number of unvisited nodes by the end of the simulation. This likely results from maintaining an equal proportion of initial spreaders across all networks. We compare between three different real world networks- a. Enron Email Network: Enron email communication network covers all the email communication within a dataset of around half million emails. Nodes of the network are email addresses and if an address i sent at least one email to address j, the graph contains an underacted edge from i to j. It was originally released by William Cohen at CMU. It contains 36692 nodes and we have set initial number of spreaders as 100 for both information types A and B (total 200 spreaders). b. Facebook Combined Network: This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using the app and the users were anonymized to protect their identities. We used the combined edge list for our simulation which consists of all the friend lists. There are 4039 nodes and the number of initial spreaders is same as enron network. c. Wikipedia Network: The data was collected from the English Wikipedia (December 2018). These datasets represent page-page networks on specific topics (chameleons, crocodiles and squirrels). Nodes represent articles and edges are mutual links between them. We have used the "crocodile" network for our simulation. There are 11631 nodes and initial spreaders are same as before.

We can see from fig 5(E) that because of the relatively less number of nodes w.r.t. spreaders and also the topology, facebook network has the highest proportion of spreaders at its peak. Also it reaches that peak faster than the other two networks suggesting that the relative number of spreaders to the total participants in the network has positive associativity with the amount of information spread. Interesting thing to note though, is that the Wikipedia network manages to converge the quickest even after having almost triple the nodes as compared to the Facebook network. This might be because the Wikipedia network never managed to reach an equal proportion of spreaders w.r.t the Facebook network. The largest network of these three, Enron, took the longest to converge. Thus, information stayed in the network for the longest but was being spread by a very small proportion of participants. Figure 5(F) shows that the Facebook network has the highest proportion of participants switching from spreading one type of information to another. A very small proportion of nodes were switching information in the Enron network, leading to the conclusion that most nodes were either ignorant or just not switching. FIgure 5(G) follows the same trend as Figure 5(E), simply because the process of becoming stiflers is a probabilistic process. Also, the number of stiflers is proportional to the number of spreaders. Hence, Facebook network has the highest proportion of stiflers followed by the Wikipedia network and the Enron network. In Figure 5(H) we can see that almost 50% of the nodes in the Enron network never received any kind of information, showing how inefficient information diffusion is in sparse networks. Wikipedia network follows right after, having about 20% its nodes as unvisited on average in the stable state. Almost all the nodes received some information in the Facebook network (being the most dense), as it had only 5% if its nodes unvisited.

V. Conclusion and Future Work

We analyze the spread of two conflicting pieces of information within a social network. We began by adapting the SIR model to include various human tendencies, such as the willingness to adopt information (Conservative, Innovative, Ordinary) and personal biases. We modified the link strengths in the network graph to represent the trust levels between two network participants. Also, we conducted simulations across various large networks, including the Facebook-Combined Network and the Enron Email Network. The results are consistent across all graphs and in accordance with general social hypotheses. Further, we could enhance the realism of the network by making it dynamic, altering node characteristics during simulations, and modeling the nodes' probabilistic properties based on different distributions.

Acknowledgment

This work was partially supported by the Doctoral fellowship in India (DIA) programme of the Ministry of Education, Government of India.

References

[1] Esteban Ortiz-Ospina (2019) - "The rise of social media" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/rise-of-social-media' [Online Resource].
[2] Yaqub, M.Z.; Alsabban, A. Knowledge Sharing through Social Media Platforms in the Silicon Age. Sustainability 2023, 15, 6765. https://doi.org/10.3390/su15086765.
[3] Soga, K.; Yoshida, S.; Muneyasu, M. Graph-Based Interpretability for Fake News Detection through Topic- and Propagation-Aware Visualization. Computation 2024, 12, 82. https://doi.org/10.3390/computation12040082.
[4] Nguyen, Le, "A Graph-Based Approach to Studying the Spread of Radical Online Sentiment" (2023). Thesis. Rochester Institute of Technology.
[5] Kermack, W. O., and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A, 115(772), 700-721.
[6] Kumar, P., Sinha, A. Information diffusion modeling and analysis for socially interacting networks. Soc. Netw. Anal. Min. 11, 11 (2021). https://doi.org/10.1007/s13278-020-00719-7.
[7] D. Easley and J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge: Cambridge University Press, 2010.
[8] Fu, Guiyuan, et al. Analysis of Competitive Information Diffusion in a Group-Based Population over Social Networks, 1 July 2019 https://doi.org/10.1016/j.physa.2019.03.035.
[9] da Silva PCV, Velásquez-Rojas F, Connaughton C, Vazquez F, Moreno Y, Rodrigues FA. Epidemic spreading with awareness and different timescales in multiplex networks. Phys Rev E. 2019 Sep;100(3-1):032313. doi: 10.1103/PhysRevE.100.032313. PMID: 31640001; PMCID: PMC7217501.