Social Media Impact
Misinformation Flow on Social Media

Progress Report 1

October 10th, 2023

Sagar Nandeshwar (260920948)

Description

The primary objective of this project is to study the information flow on social media platforms such as twitter or facebook and how it influences users' behavior. In order to investigate them, I will develop a network framework, an information flow mechanism and an ML model.

Overview:

I have divided the project into 4 sections, which are as follows:

1. Framework development:

I will build a graph-based framework to understand the underlying structure of social media platforms and will test different parameters to best analog Twitter.
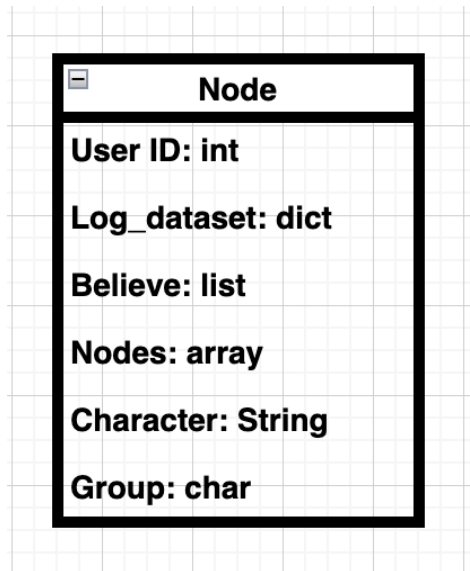
Proposed Model: Let the graph G(V,E) represent the framework of twitter, where
V = set of nodes that represent users or sub-group of people with similar believes
E = set of edges that represent their connection in the form of followers or subscription
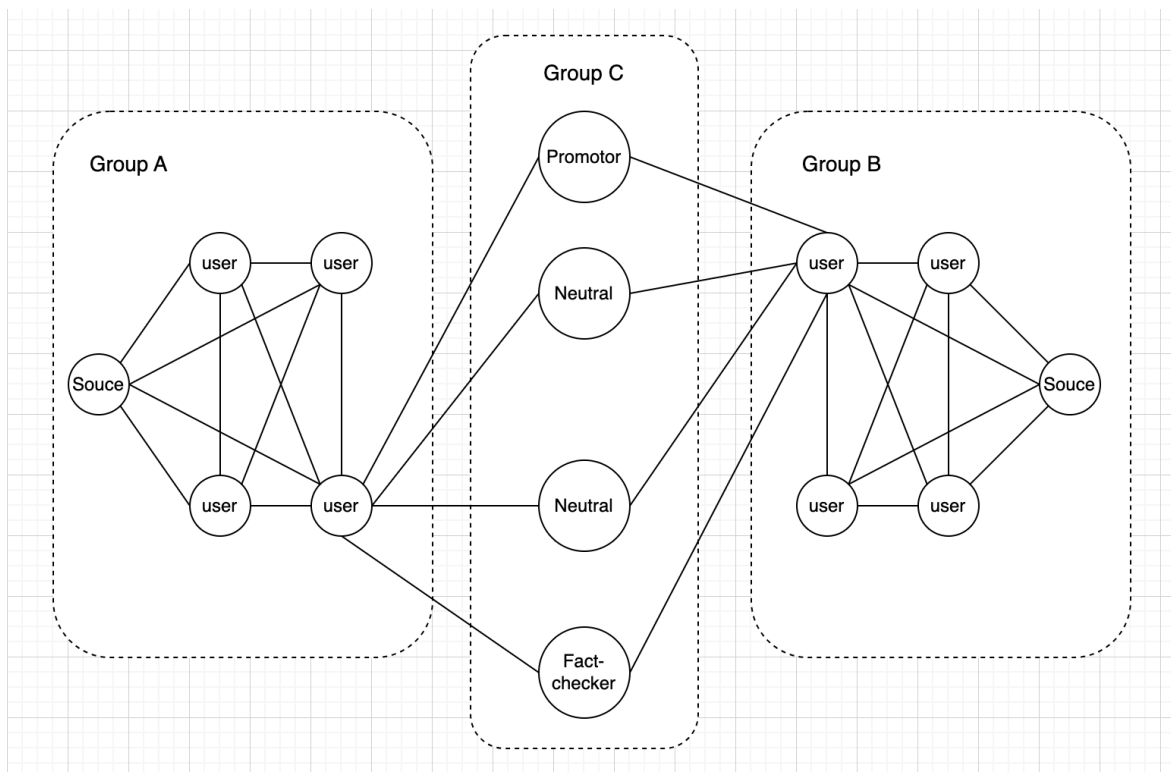
Each node will be as follows:

2. Information Flow Mechanism:

To study and implement the information flow I will modify the Bootstrap Percolation[1] model for the initial set-up and then the First Pass Percolation[2] model for the information flow in multiple cycles.

3. Developing ML model

Once the network has been infiltrated with information, I will use an NLP based non-linear ML model to predict the user's behavior. This model will be trained on supervised datasets consisting of lists of facts and fake facts. It will take nodes' log dataset as input, vectorize it and produce an output which will be the user's beliefs on the selected topic.

4. Analogy with twitter:

I will collect up to 40000 tweets on 3 different topics and will categorize them into facts and fake facts. This dataset will be gradually introduced to the graph from the two sources nodes. After enough flow cycles, I will use the nodes dataset, consist of the same tweets, and predict the user's final behavior.

End of Semester Statement

By the end of the term, I should have a framework to analog twitter, a working information flow mechanism, a trained ML model to predict users' behavior, and lastly users' predictions and an analysis of change in their behaviour.

Deliverables Table

Developer	Due Date	Task	Points
Sagar Nandeshwar	Oct 15th, '23	Framework Implementation	2
Sagar Nandeshwar	Nov 1th, '23	Information Flow Mechanism	2
Sagar Nandeshwar	Nov 30th, '23	ML model and Twitter data collection	2
Sagar Nandeshwar	Dec 5th, '23	Observation and analyses	2

Experiment
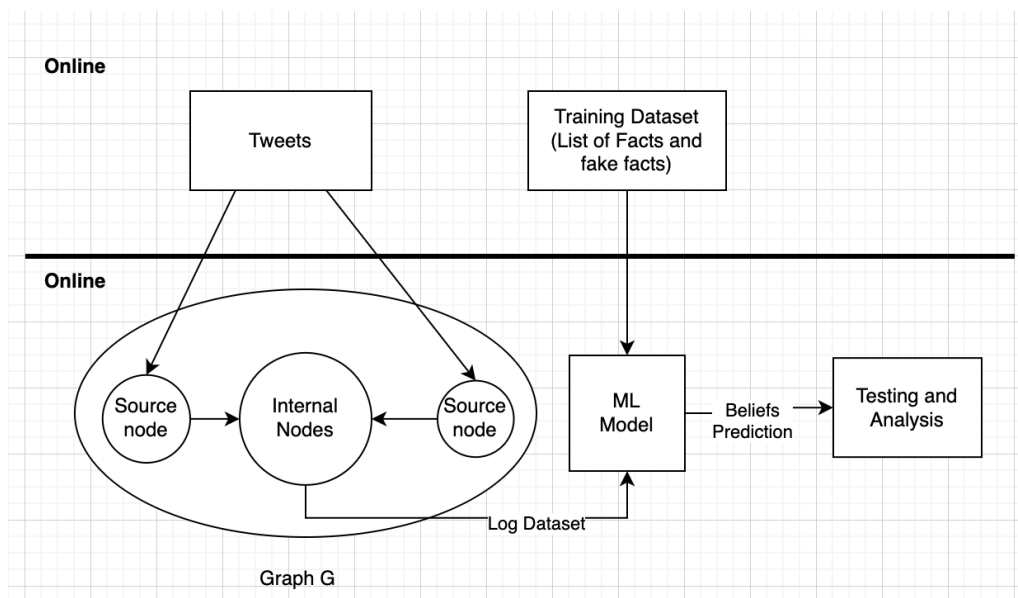
I divided the nodes into three subgroups, say A, B and C. Group A will believe in the fake facts, Group B will believe in real facts and Group C will be neutral. There will be two source nodes, one in group A and one in group B, which will spread the information in their respective groups.

Now, group A and group B are connected through group C, which will be filled with three kinds of nodes, namely promoters, fact-checkers and neutral nodes. The promoter node will receive a data packet from one group and will pass to another with some degree of exaggeration. The fact-checker node will read the data packet from one of the groups and will pass or destroy it depending on the data validity, lastly the neutral nodes will simply pass the data packet from one group to another.

Each of the nodes will log the information they receive in a dataset. I will use these datasets to predict the nodes' beliefs with the help of an ML model, once the network has been infiltrated with information.

Lastly, I will analyze the change in users' behaviors depending on the ML models results for each node, which would give some insight on the effect of misinformation on users.

I will also test and study the effects of the following
- Network Framework parameters
- Flow mechanism parameters (Bayesian parameter estimation and maximum likelihood)
- ML model parameters (Text Pre-processing, Hyperparameters, and K-fold training etc)
- Effect Promoter, Fact-checker, and Neutral nodes

Expected Results

I expect to see change in behavior of some nodes, specifically the one who are farthest from the sources as they can be infiltrated with more misinformation.

Technology Stack of the Application

- I will build the framework project in Java using Java Socket Programming
- I will build the ML model in Python using scikit-learn library and NLTK
- I will be using Twitter's API to collect approximately 40,000 tweets for testing.

Repository Description

I've created a new repository Social_Media_Impact (https://github.com/Sagarnandeshwar/Social_Media_Impact), and will soon update it with the framework and project description.

Existing Repo Access (if applicable)

	Yes
	No

Slack Access

Yes	Yes
	No

Lab Google Sheet Access

Yes	Yes
	No

Updated Yellow Sections of Lab Google Sheet

	Yes
No	No

Previous Studies

The project been inspired from the following papers

1. "The spreading of misinformation online" - Michela Del Vicario, Alessandro Bessi, Fabiana Zollo et al
2. "Impact of correcting misinformation on social disruption" - Ryusuke Iizuka et al
3. "Party Prediction for Twitter" - Reihaneh Rabbany et al
4. "Fast and Attributed Change Detection on Dynamic Graphs with Density of States" - Reihaneh Rabbany et al
5. "Information Spread on Social Networks" - Joseph Vybihal and Simon Overgaard
6. "Networks: Social" - F.N. Stokman

Paper 1 inspired the original motivation for this project, however my project will be more focused on implementing Network Framework and spread mechanism than just analysing the data flow. Paper 3 and 4 inspired the ML model for user's prediction and dynamic graph design, respectively. However they are not directly used in the context of Social Media platforms, and in this project I will also try to unify them. Paper 2 shows the effects of fact-checkers, which I plan to include in the graph network. Lastly, paper 6 gives a theoretical perspective on some important elements of Network Science.

Reference
[1] "K2,t,-Bootstrap Percolation" by Mohammadreza Bidgoli Mohammadreza Bidgoli, Ali Mohammadian and Behruz Tayfeh-Rezaie
[2] "First-Passage Percolation" by Harry Kesten