Department of Computer Science

CS 429/529 – Dynamic and Social Network Analysis

Project Short Proposal

Doruk ALTAN - 21401362

Alperen CAN - 21601740

Instructor: Miray KAŞ

Teaching Assistant: Gözde YAZICI

Bilkent University

24.10.2022 (dd/mm/yy)

Our project aims to do a network analysis of Covid vaccine misinformation. We will study twitter data. Specifically, tweets that contain the words "*Coronavirus vaccine*" and "*Covid 19 vaccine*". We plan to gather tweets from the Twitter API to serve as our dataset. We then aim to detect malicious networks that spread misinformation. It is important because social network analysis of the users who spread misinformation is helpful for authorities to identify the people having an influencer role, thus facilitating the assessment of the scale of misinformation. In order to classify the tweets, we will make use of BERT which is a pre-trained language model to help solve Natural Language Processing (NLP) tasks such as sentiment analysis.

Our main goal is to extract features from tweets via BERT and use several classifiers on this data and compare them such as Random Forest, Support Vector Machines (SVM). From this network we seek to gain insight about the users that have the most influence in the spread of misinformation surrounding the Covid 19 vaccine. We submit that the most influential users will be those that have many likes, retweets, etc… and those that are intermediary between many nodes. Thus, we decided to use the metrics of betweenness centrality and degree centrality. We believe these metrics to be good indicators as likes and retweets are shown as connections between nodes and nodes with the highest number of connections rank highest in degree centrality. To identify the intermediary nodes, on the other hand, we will analyze the betweenness centrality. Depending on our progress throughout the semester, we also hope to fine tune BERT to get better feature extraction and better classifier results.

Over the course of the project, most challenges we expect to face are getting familiar with new software and technologies. When it comes to NLP, semantic analysis or Machine Learning, we are both novices. However, there are many resources, tools, models and tutorials that we can use to overcome our lack of expertise with such subjects. For this reason, further exploration and improvement on the BERT training is left as a potential extra feature. In this project we aim to highlight the analysis of the network and how different metrics and classifiers compare to one another rather than the performance of the classifiers or ML techniques.