Conspiracy Theory Dynamics in Online Social Networks through Unsupervised Learning Models

The 5G Corona Conspiracy

Master Thesis: Kaspara S. Gåsvær

Start: August 2021 - End: 2022

This thesis aims to better understand conspiracies theories and their diffusion in online social networks using physics inspired methods for modeling. This will be done with focus on boltzmann machines and statistical mechanics among other unsupervised learning methods. Using the example of the 5G conspiracy and the resulting Wico-Text and Wico-Graph datasets, we will develop an unsupervised learning approach based on spreading graphs, tweets, and twitter-user features.

Introduction

About the incident

Soon after the COVID-19 outbreak in Wuhan, China, a series of tweets surfaced on Twitter containing insinuations of a possible link between the virus and 5G wireless technology. The first tweet was posted already before the virus spread from its place of origin but seemed initially to gain very little attraction. It would take many weeks before the full scope of the problem showed itself as a series of arson attacks on 5G towers in multiple countries, including the UK. Such spreading of misinformation online that leads to real-world implications can be classified as what is known as *digital wildfires* and has been ranked as one of the top global risks in the 21 century by the World Economic Forum. When posing real-world issues of this magnitude one can begin to understand why the knowledge of how inaccurate or misleading information spreads online is a topic worth while investigating.

Datasets

This thesis's source of data is a subset of tweets collected between January and May 2020. More specifically, tweets containing keywords related to the COVID-19 pandemic **and** 5G, as well as a set of corresponding spreading graphs.

- **Wico-Text:** a Labeled Dataset of Conspiracy Theory and 5G-Corona

- **Wico-Graph:** a Labeled Dataset of Twitter Subgraphs based on Conspiracy Theory and 5G-Corona Misinformation Tweets

The **Wico-Text** dataset is a misinformation dataset connected to the 5G-Corona conspiracy that contains 10,078 tweets from Twitter's follower network hand-labeled into four classes: *promoting 5G conspiracies*, *promoting other conspiracies*, *not promoting conspiracies,* and *undecidable.* They are referred to as tweets but contain both tweets and a small number of *quotes* and *retweets*.

The **Wico-Graph** is a dataset that contains extracted subgraphs from 3000 manually classified Tweets from the Wico-Text dataset. All 3000 subgraphs are classified into three categories: *subgraphs of tweets that spread misinformation about the 5G corona conspiracy, subgraphs of tweets that spread other types of conspiracies,* and *subgraphs of tweets that spread neither.* It is important to know that there are several ways of sharing information on Twitter

1. Tweet - A new post
2. Reply - A comment to a tweet
3. Retweet - Sharing an already existing tweet or reply from someone else
4. Quote - A retweet with either additional comments or modifications from the original

Each subgraph was induced by the Retweets of each Tweet among the 3000 hand-labeled ones (see Wico-text for more detail about the labeling). The Replies are included in the graph of the Tweet it commented on as well as treated as the first Tweet in a new subgraph.

Time schedule

We will section the thesis in roughly 4 stages:

1. **Combining/finding features of graph and text/tweets**
2. **Introduce basic Unsupervised Clustering methods**
3. **Move on to more complex clustering methods and other experimental methods like**
    1. **Restricted Boltzmann Machines (RBMs)**
    2. **T-distributed stochastic neighbor embedding**
    3. **Possible other methods of interest.**
4. **Simulation of spreading over time**

Sections 1 and 2 will be done during the Fall semester 2021 while sections 3 and 4 belongs to the Spring semester 2022.

Possible areas of interest

- **Distance metric learning**: k-means clustering (divide *n* observations into *k* clusters). Preliminary cluster partition, define proximity measures (Euclidean?). Hierarchical clustering Algorithms?

- Investigate **clustering coefficient** (the tendency of clustering of the nodes in a graph)

- **Sentiment score:** (polarity in text, negative or positive emotion/opinion). Look into how the emotional state of a person (by the sentiment score of a tweet) affects the probability of them spreading misinformation / the rate of which they spread misinformation (posting rate).

- **Simulation over time:** Can we simulate how misinformation spreads over time within a cluster/community or to other groups? Combine graphs with

information gained about crucial features? Look at how misinformation in the WISCO graph has spread -> predicting spreading patterns?

- **Bots:** Repeat experiment with and without bots

Main Supervisor:

Morten Hjorth-Jensen

Student:

Kaspara Skovli Gåsvær

Dato: 27.11.20