Project Proposal – Capstone 3: Network Analysis of r/wallstreetbets

Problem Statement

The rise to prominence of r/wallstreetbets following the Gamestop saga has been meteoric. Over the course of January 2021, the value of the stock rose from approximately $17 at the beginning of the month to $347 on Jan. 27th. This was largely fueled by the mania that permeated the reddit community, and while the GME circumstances were exceptional, it has been established that intelligently analyzing the trends of r/wallstreetbets can potentially prove fruitful for making investment choices.

The goal of this project is to apply social network analytic methods to determine 'hot' stocks according to the subreddit. I will analyze my findings in relation to stock performance.

Why this Study is Important

Reddit is one of the largest social media sites in the world, and features millions of posts on just about any topic of potential value. Analyzing the network structure of r/wallstreetbets will provide insight into how positive and negative sentiment towards specific stocks is propagated, information which can potentially help inform good investments.

Value to Extract

 The goal of this study is to develop a network representing the wallstreetbets subreddit's attitude towards stocks of interest. Through various data science techniques, I will extract meaningful information from this network, and how the attitude of the subreddit towards certain stocks evolves.

Data we need

My intention is to extract all relevant posts on r/wallstreetbets relating to a provided list of stock ticker symbols. I will include a requirement of a minimum number of responses, since we are concerned largely with community interaction. I will use the Reddit PushShift API to accomplish this.

Methods Employed

The code for this project will be in the form of Python Jupyter Notebooks. As stated above, I will be using the Reddit API to extract the necessary data. I will be using the network Python library to construct and analyze the graph structure of the data. It is possible that I will also include sentiment analysis techniques in this project.