Evaluating Differences in Public Sentiment Surrounding the Nvidia Stock Across Politically-Divergent Social Media Platforms

Srivatsa Kundurthy

Introduction

How does public sentiment surrounding "Magnificent 7" technology stocks such as Nvidia differ across mainstream and right-wing social media platforms?

From a financial perspective, the "Magnificent 7" tech stocks, which include Microsoft, Apple, Nvidia, Alphabet, Amazon, Meta, and Tesla, have come to dominate market speculation in the past decade. These corporations have all achieved a market capitalization greater than 1 trillion USD, and are major players in the next-generation of influential technologies (Caulfield, 2024).

The growth of the Nvidia stock in 2024 is especially notable: driven by investors' bullish confidence in AI, the chip manufacturer has seen a 94% increase in share prices from January through May of the year. Magnificent 7 stocks are credited with driving a significant portion of the market's returns in recent years: hence, it is incredibly fruitful for investors to be able to analyze and predict the behavior of these stocks.

In the digital era, social media discourse has emerged as a new big-data dimension in stock market analytics (Bouadjenek et al., 2022). Traditionally, stock market modeling has been accomplished through observation of financial markers such as P/E ratio, market capitalization, return ratios, and equity ratios (Fataliyev, 2021). While these proxy measurements are purposeful, stock market motion is most directly a function of public sentiment: if individuals feel positively about a stock, they buy it, and if they feel negatively about a stock, they sell it (Shah, 2024). Because standard surveying methods are not realtime, and do not scale effectively to capture the unbiased opinions of investor populations, public sentiment has been challenging to quantify directly. However, the advent of big data techniques and the ample availability of real-time thoughts and opinions on social media platforms has refreshed our ability to predict the stock market based on the aggregate opinions of investors. In one notable study, Bollen et al. (2011) analyzed over 9.8 million tweets and found that changes in market-related discourse on Twitter preceded subsequent movements in the Dow Jones Industrial Average by up to four days, suggesting a predictive relationship between social media sentiment and stock market behavior.

Given the popularity of mega-cap tech stocks, it is likely that social media chatter surrounding these companies is rich and conducive for automated NLP methods, making the public sentiment reflected by social media data a valuable tool for capturing the motion of some of today's biggest market movers.

However, when we use social media to analyze a stock, we make the critical assumption that the user bases of the social media platforms are representative of the true population we wish to study–that is, we assume that social media platforms sample uniformly from the entire population that we are interested in. It has become increasingly difficult to find social media platforms that capture the general population, as user bases on social media have been shifting rapidly in recent years. In 2020, mainstream platforms like Facebook, Instagram, and Twitter started rigorously enforcing content moderation policies in order to crack down on COVID/election misinformation (Chen et al., 2021). This was a push factor for right-wing groups, and they began to voluntarily leave mainstream social media sites in favor of apps like Gab, Telegram, and Truth Social, which promise to "uphold free speech" and limit content moderation. These platforms have seen significant growth: a recent report found that 56% of Americans had heard of at least one alt-tech platform. Moreover, prominent accounts on these platforms often appeal to right-leaning or pro-Trump values, with 54% of accounts of alt-right platforms explicitly endorsing conservative values (Stocking, 2022).

This trend of "conservative drain" from mainstream social media limits the breadth of financial conclusions that can be drawn from sentiment analysis of social media data from a singular platform. For example, a platform like Truth Social would have users who hold predominantly conservative investment perspectives, such as a preference for defense stocks (Rojas, 2021). Therefore, predicting defense stock prices based off of Truth Social data alone will yield a biased and inaccurate stock model.

The research question we study has important implications. It focuses on two significant, opposing forces that are affecting financial analytics and social media platforms. Firstly, big data methods have increasingly led researchers to predict stock behavior by using social media to measure public sentiment. At the same time, however, social media platforms are going through a major population drift in which conservative users are migrating to new spaces with limited content moderation. This violates the critical assumption made by financial researchers, which is that social media platforms sample uniformly from a general population of investors. In this work, we will conduct a case study of the burgeoning Nvidia stock, studying the sentiment of posts that mention the company across mainstream and alt-right platforms. In our analysis, we seek to understand users' sentiment towards Nvidia and the context in which Nvidia is mentioned on each platform. We ultimately intend to draw conclusions on whether or not our data suggests that the user bases across mainstream and alt-right platforms sample from the same distribution. If the evidence suggests that these platforms represent skewed facets of the population, then there are new considerations that must be made when using social media platforms to model financial markets. All code is available in the GitHub repository.

Data Sources

This project requires time-stamped post data from a mainstream social media platform and time-stamped post data from an alt-right social media platform. Because the project is centered around sentiment analysis, the platforms from which post data is collected should be "text first", meaning that text makes up the primary mode of expression (a non-example is Instagram, which is primarily for photo/video sharing).

Mainstream Social Media Platform

For mainstream social media data, this project uses X, formerly Twitter. X is an immensely popular social media platform that centers around real-time expression of opinions. It can be searched by hashtag and keyword, meaning that posts regarding specific stock tickers and company names can be scraped. Additionally, tweets are time-stamped, so posts can be aligned directly with major stock motion. The rich textual data that X contains is highly conducive to sentiment analysis.

Alt-Right Social Media Platform

For alt-right social media data, this project uses Gab. Gab is a social media platform that is known for its far-right user base. Though significantly less popular than X, Gab has an almost identical layout and purpose, allowing for rapid sharing of opinions and thoughts. With very few restrictions and limited moderation, the platform caters heavily towards alt-right groups. It can be indexed according to hashtag and keyword, meaning that specific stock tickers and company names can be scraped. Additionally, posts are time-stamped, so data can be compared directly with major stock motion. The rich textual data that Gab contains is highly conducive to sentiment analysis.

Data Collection & Processing

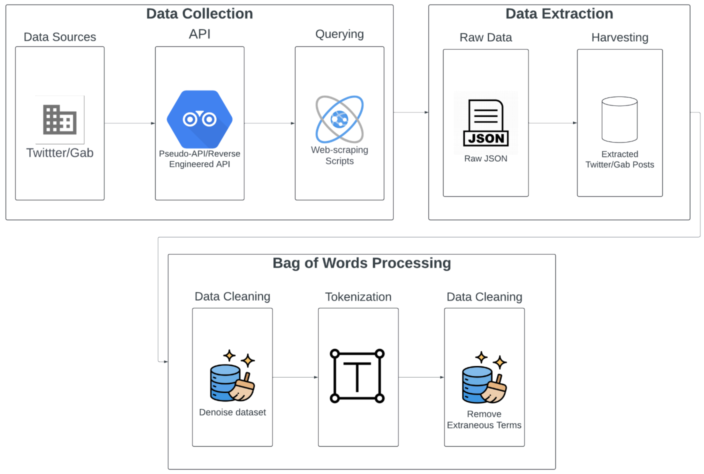

Figure 1: A Diagram of the Process of Data Collection & Processing

Data Collection

Posts mentioning “nvidia” are scraped from X and Gab over the range February 1, 2024 to February 29, 2024.

“Twscrape”, a private X API scraper, is used to collect data from X. One core benefit of “twscrape” is that it allows for automatic account switching, which prevents loss of progress if a scraping bot is banned by switching scraping seamlessly to a different X account. In total, 306 posts are collected from X. While this is non-exhaustive, it is a representative sample over the date range.

Data collection from Gab proved to be more challenging, as the alt-right platform does not maintain a user-facing API. Hence, we reverse-engineered the site’s private API. To accomplish this, we utilized tools in the Postman app to inspect the HTTP requests and responses made when keyword searches were made on the Gab.com website. By examining the request URLs, headers, and payloads, we identified the query parameters and the structure of the API requests, and constructed a template for access to the Gab API.

Subsequently, we wrote a simple Python script, using the “requests” library to automate the data scraping process. Our script made HTTP requests to the reverse-engineered API endpoints and handled pagination to traverse through the search results. Throughout this process, we ensured that our methods adhered to ethical scraping guidelines, preventing any undue strain on Gab servers. In total, we collected 312 posts from Gab.

Data Extraction

The data collected via webscraping takes the form of raw, semi-structured JSON files representing posts on each platform. These JSON files are parsed for the text field, and these text fields are dumped into a singular text file.

Bag of Words Analysis

Finally, standard bag-of-words techniques are applied to process the text data. The data is pre-processed through the removal of noise such as non-unicode characters and newline characters, then tokenized, and then processed again to eliminate stop words and capitalization, leaving a vector of words that can be efficiently analyzed.

Visualizations

We generate several visualizations of our data in order to draw conclusions about differences in sentiments regarding “Nvidia” across politically-opposing social media platforms.

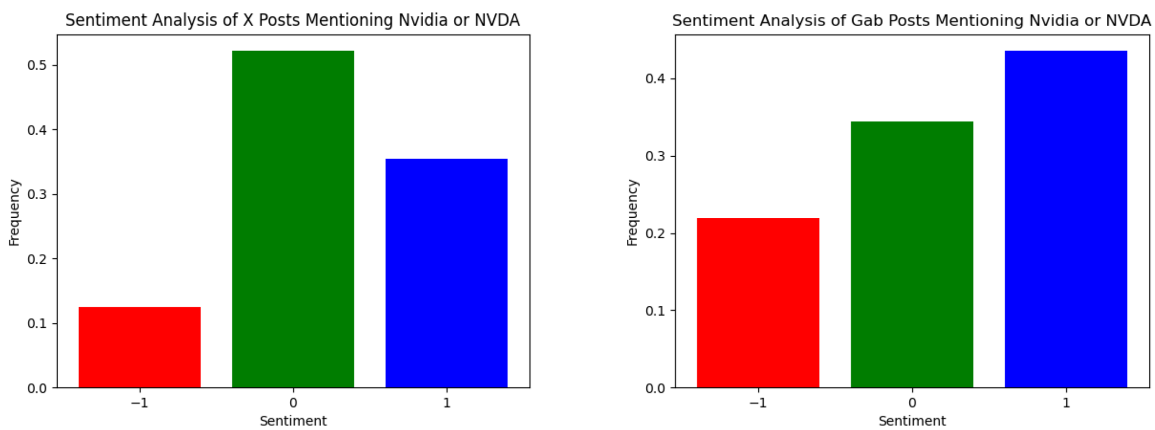

Figure 2: Frequency of Positive, Neutral, and Negative Tweets and Gabs Mentioning Nvidia

In Figure 2, the nltk package is used to compute sentiments for each Tweet in the file of Tweets, and each Gab post in the file of Gab posts. “-1” denotes negative sentiment, “0” denotes neutral sentiment, and “+1” denotes positive sentiment.

In the following figures, the visuals characterize the topics and nature of the posts for each category of sentiment across each platform. “Figure Xa” will be for Twitter data, and will include a topic network with seed word “Nvidia” on the left, and a plot of word frequency of the top 10 words on the right. “Figure Xb” will be the same, but for Gab data. It is possible to zoom in and see the text in the topic networks; alternatively, the visuals are available with the GitHub repository for this project.

Negative Posts.

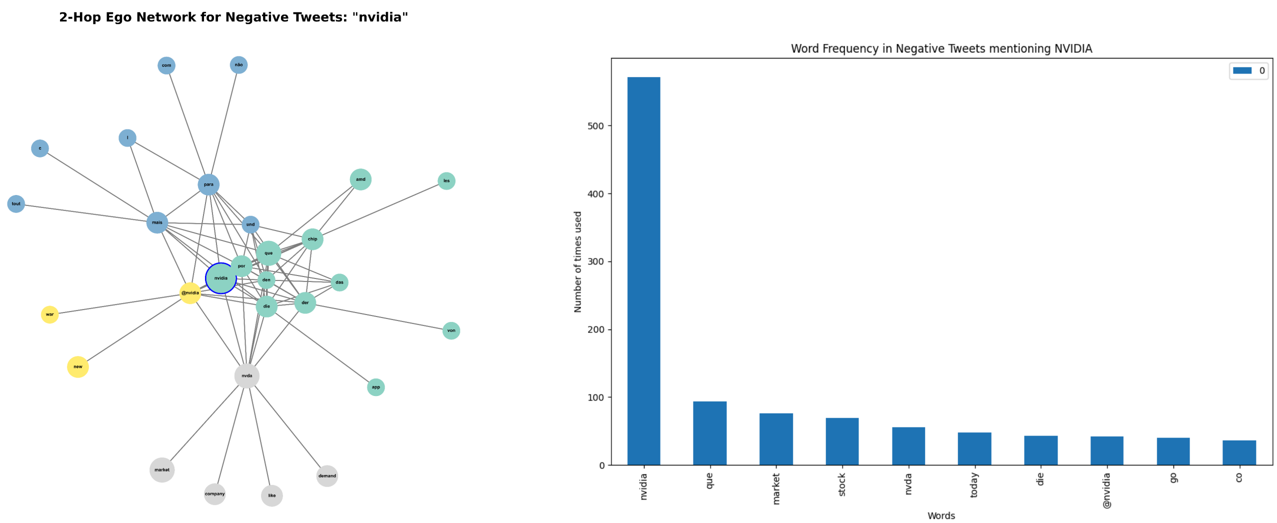

Figure 3a: Common Topics & Frequent Terms in Negative Tweets about Nvidia

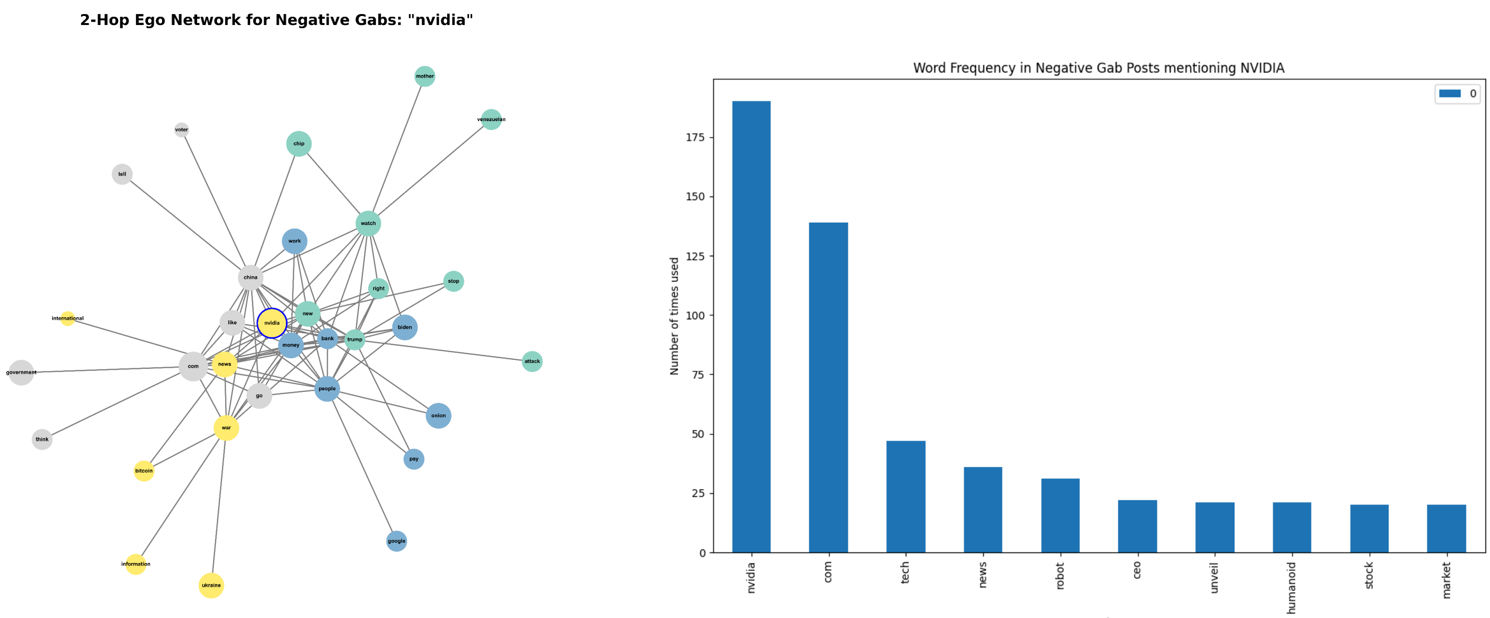

Figure 3b: Common Topics & Frequent Terms in Negative Gab Posts about Nvidia

Neutral Posts.

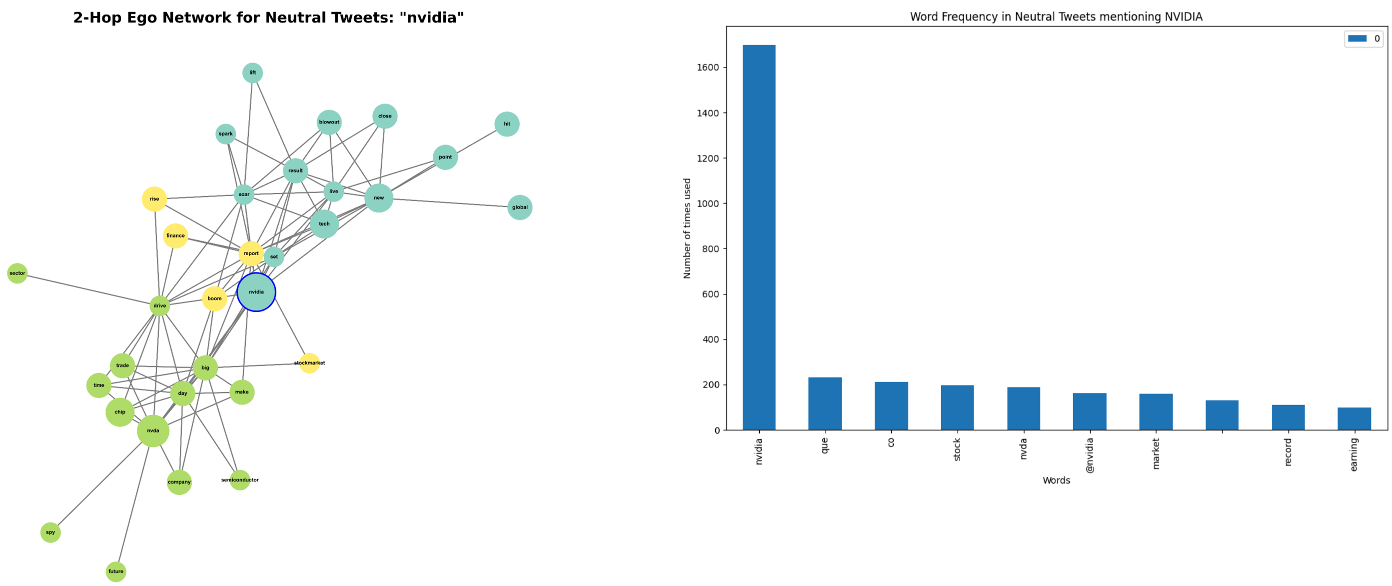

Figure 4a: Common Topics & Frequent Terms in Neutral Tweets about Nvidia

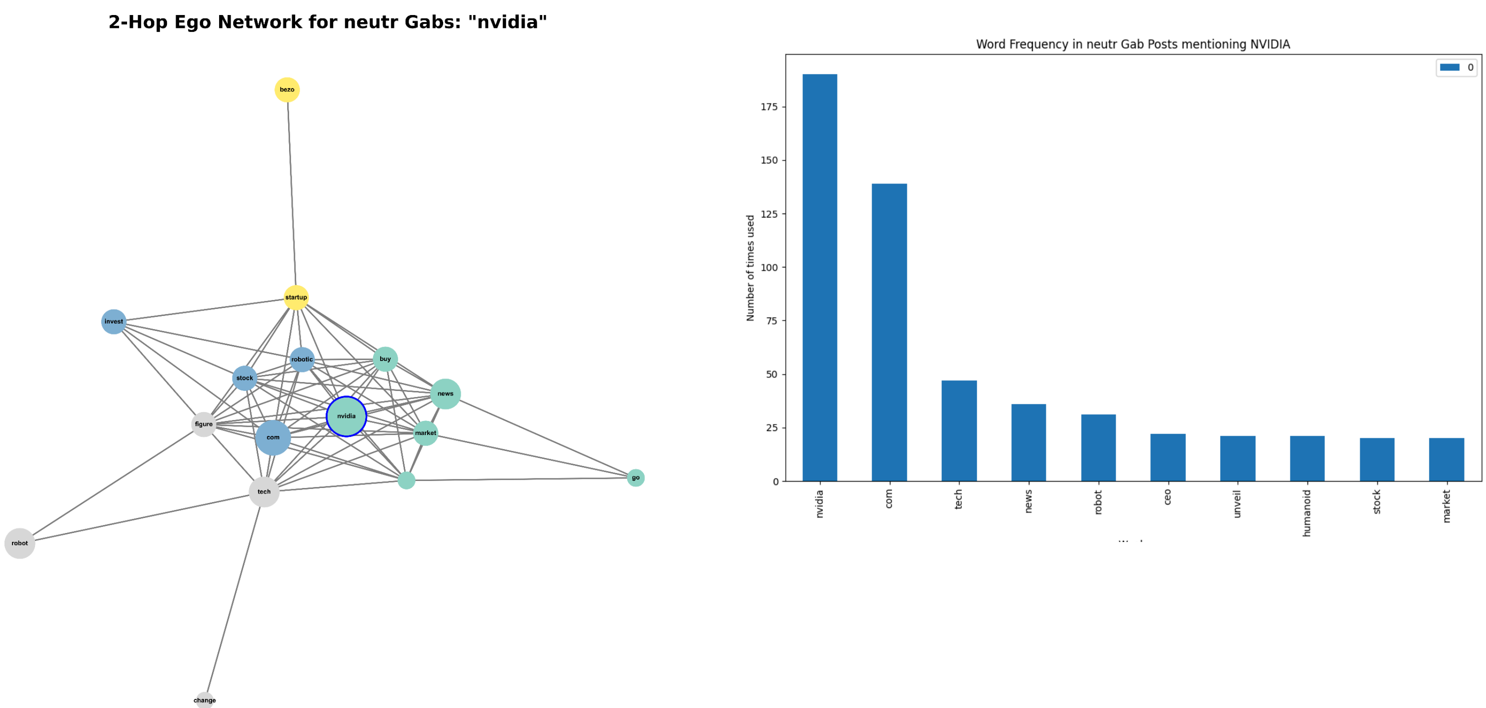

Figure 4b: Common Topics & Frequent Terms in Neutral Gab Posts about Nvidia

Positive Posts.

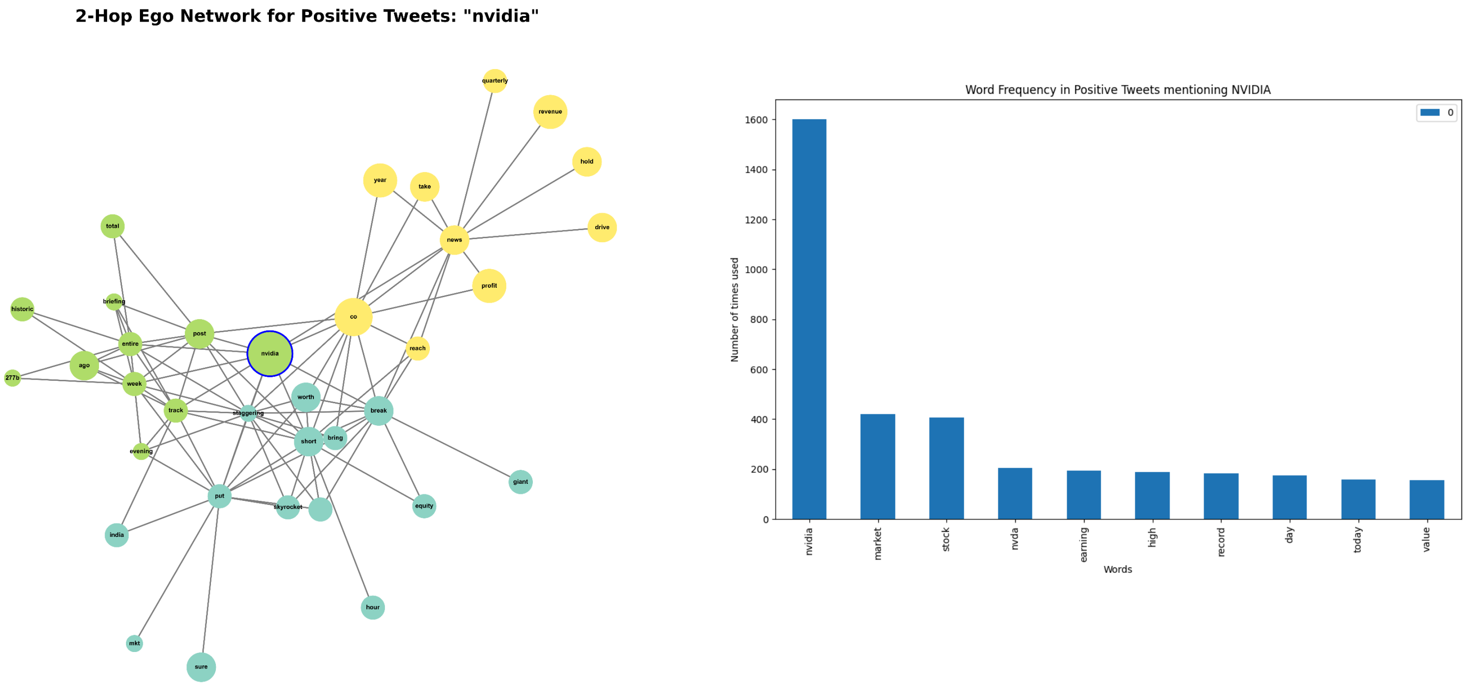

Figure 5a: Common Topics & Frequent Terms in Positive Tweets about Nvidia

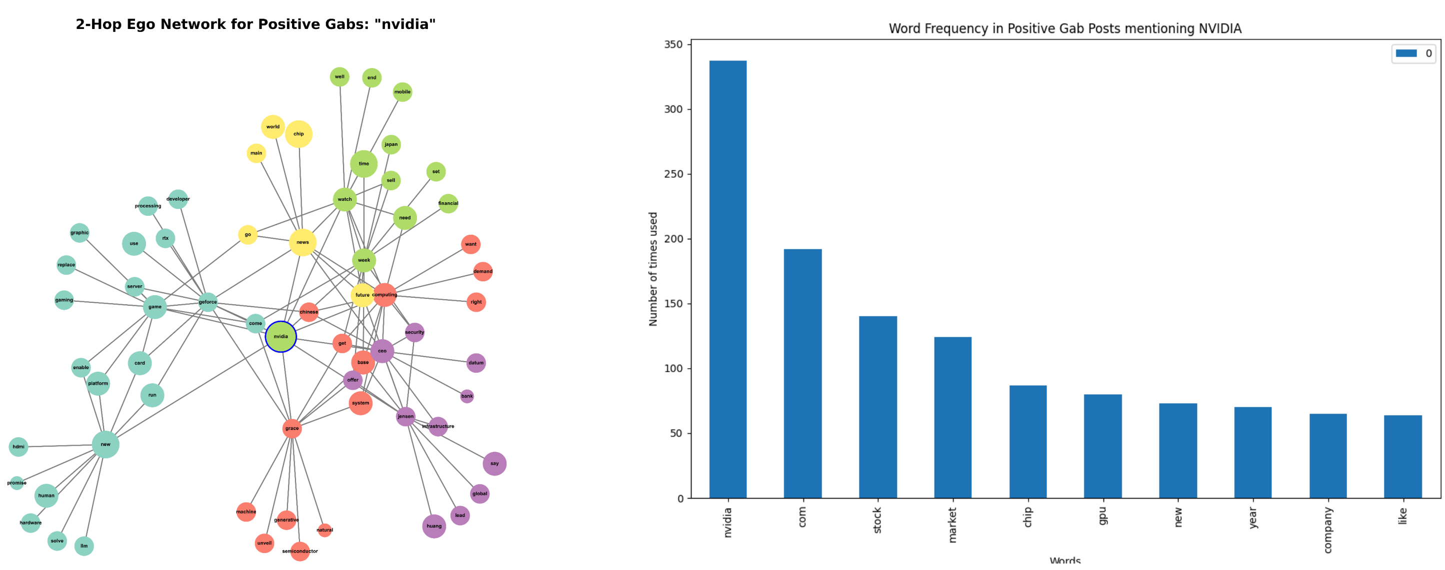

Figure 5b: Common Topics & Frequent Terms in Positive Gab Posts about Nvidia

Analysis & Discussion

We evaluate our research question in accordance with these visualizations.

The first visualization we developed is Figure 2, which shows that, of the X posts we collected, approximately 12% were negative, 32% were positive, and over 50% were neutral. On the other hand, of the Gab data that we collected, approximately 21% of posts were positive and 44% of posts were negative, while only about 35% were neutral. This result suggests that Gab posts discussing Nvidia in the month of February are more polarized than Tweets over the same duration–that is, they align more with a definitive positive or negative scoring than a neutral characterization. This outcome is consistent with literature that describes Gab as an echo chamber for radical views (Weigel, 2024).

We initially speculated that a greater degree of Gab posts would contain negative sentiment discourse, as the conservative viewpoint towards big tech stocks has been jaded at best. Following observation of this sense of polarity on the Gab platform, we were motivated to understand the specific topics and keywords that were prevalent in each sentiment category on both platforms.

For negative sentiment posts, Figure 3a for X data unfortunately returns non-English noise. X is notoriously subject to bot activity, and it appears that the nltk sentiment analysis model characterized most non-English terms as being of negative sentiment. Due to a lack of funding, we were unable to pay for the API service that would detect and remove non-English strings from our corpus, so our findings for negative-sentiment tweets are inconclusive. However, the topic model in Figure 3b for Gab posts shows that potentially inflammatory topics such as “Biden”, “China”, “Trump”, “war”, “Ukraine”, and “bitcoin” are frequently discussed in the context of Nvidia. Gab is well known for perpetuating misinformation, and further research may uncover theories linking Nvidia with common conspiracy narratives.

For neutral posts, Figure 4a for X data reveals a dense network of common terminology used to describe stock growth, including “boom”, “blowout”, “rise”, as well as business-related terms such as “semiconductor”. On the other hand, Figure 4b for Gab data is remarkably sparse in comparison, containing only a handful of loosely related topics like “buy” and “market”: it appears as though there are very few cohesive comments on Gab that mention Nvidia in a neutral manner. This stark contrast in topics for neutral posts once again reinforces the polarization of discourse on the Gab platform.

For positive posts, Figure 5a for X data is a very dense network containing several common financial topics like “short” and “equity”, as well as topics like “skyrocket” and “historic” that are enthusiastic towards the growth of the stock. Interestingly, Figure 5b for Gab data is far denser, containing a breadth of topics including “computing”, “CEO”, “demand”, “hardware”, and “future”. Curiously, it appears that positive tweets mentioning Nvidia on Gab seem to engage in insightful discussion regarding Nvidia Corp as a whole. These results support the outcome that Gab is a polarized platform in comparison with mainstream platforms, in that the nature of positive discussions on the platform are also richer in detail.

Conclusion

While our analysis is primarily qualitative and our results are not statistically significant, our observations support the hypothesis that, in the complex landscape of politically-divergent media platforms, the user bases of social media platforms are skewed and do not reflect the entire investor population. Sentiments surrounding the Nvidia stock’s growth in February 2024 varied greatly across both platforms, and topics of discourse were divergent as well. Gab is highly polarized to a degree that X isn’t, and this level of polarization is reflective of the biases and opinions of Gab's underlying user base. If the platforms’ user bases were representative samples of the general investor population, we would have observed closer sentiment patterns between both platforms. Financial researchers should endeavor to include numerous, politically opposing social media platforms in their datasets, as aggregation of increasingly disjoint data across platforms will assemble a more collective representation of the general population. Otherwise, as social media platforms become increasingly polarized and concentrated, our work shows that financial researchers are losing their ability to capture unbiased public sentiment towards stocks.

Works Cited

Bollen, Johan, et al. “Twitter mood predicts the stock market.” Journal of Computational Science, vol. 2, no. 1, Mar. 2011, pp. 1–8, https://doi.org/10.1016/j.jocs.2010.12.007.

Bouadjenak, Mohamed. A User-Centric Analysis of Social Media for Stock Market ..., 2022, ssanner.github.io/papers/tweb22_userstockpred.pdf.

Chen, Emily, et al. “Covid-19 Misinformation and the 2020 U.S. Presidential Election: HKS Misinformation Review.” Misinformation Review, 10 July 2023, misinforeview.hks.harvard.edu/article/covid-19-misinformation-and-the-2020-u-s-presidential-election/.

Galen Stocking, Amy Mitchell. “3. Prominent Accounts on Alternative Social Media Sites Mostly Are Individuals, Not Organizations.” Pew Research Center, Pew Research Center, 6 Oct. 2022, www.pewresearch.org/journalism/2022/10/06/prominent-accounts-on-alternative-social-media-sites-mostly-are-individuals-not-organizations/.

Kamaladdin, Fataliyev. Text-Based Stock Market Analysis: A Review, 2021, arxiv.org/pdf/2106.12985.pdf.

Publishing, Sungarden Investment. “Mags: Charting Magnificent 7 Shows New Trend (NASDAQ:Mags).” Seeking Alpha, 29 Apr. 2024, seekingalpha.com/article/4687136-mags-charting-magnificent-7-shows-new-trend-technical-analysis.

Shah, Pratham, et al. “A comprehensive review on sentiment analysis of social/web media big data for stock market prediction.” International Journal of System Assurance Engineering and Management, 26 Jan. 2024, https://doi.org/10.1007/s13198-023-02214-6.

Warren Rojas, Camila DeChalus. “At Least 15 Lawmakers Who Shape US Defense Policy Have Investments in Military Contractors.” Business Insider, Business Insider, www.businessinsider.com/congress-members-are-trading-defense-stocks-while-shaping-military-policy-2021-12. Accessed 7 May 2024.

Weigel, Moira, and Adina Gitomer. “Hate-sharing: A case study of its prevalence and impact on Gab.” New Media & Society, 29 Apr. 2024, https://doi.org/10.1177/14614448241245349.