2021

Diffusion of Innovation

A NICHE ANALYSIS OF A NICHE CRYPTOCURRENCY

ANDREW LINDBERG

Abstract

One promising shift in the way we engage with the internet is “web3.0” – a potential future state where networks operate through decentralized protocols (rather than being tied to centralized networks and servers controlled by a select few companies). “Web3.0” is expected to be symbiotic with blockchain and cryptocurrency technologies. This has lead to a boom of investment into these technologies; wherein, picking the right technology has potential for significant capital gains.

This project seeks to gain insight into a blockchain ecosystem built on the Ethereum network called Fantom. There are many layer 1 ecosystems like Cardano, Solana, Avalanche, Polygon, and Fantom. More advanced researchers may be able to distinguish the technical advantages of one ecosystem over another. This researcher likes Fantom because of the commitment to the theme of the network by the developers on the ecosystem.

This project evaluates Twitter data, in hopes of finding a relationship between tweets and the value of cryptocurrencies related to the search term fantom.

Data

Massmine was used in conjunction with Twitter’s API to identify tweets related to the search term “fantom” from 2021-11-05 to 2021-11-14. 106,155 observations were recorded, the data was converted in R using the ndjson package into a dataframe, and filtered using the tidyverse package to 16 unique variables. Ultimately, the research used the following data points:

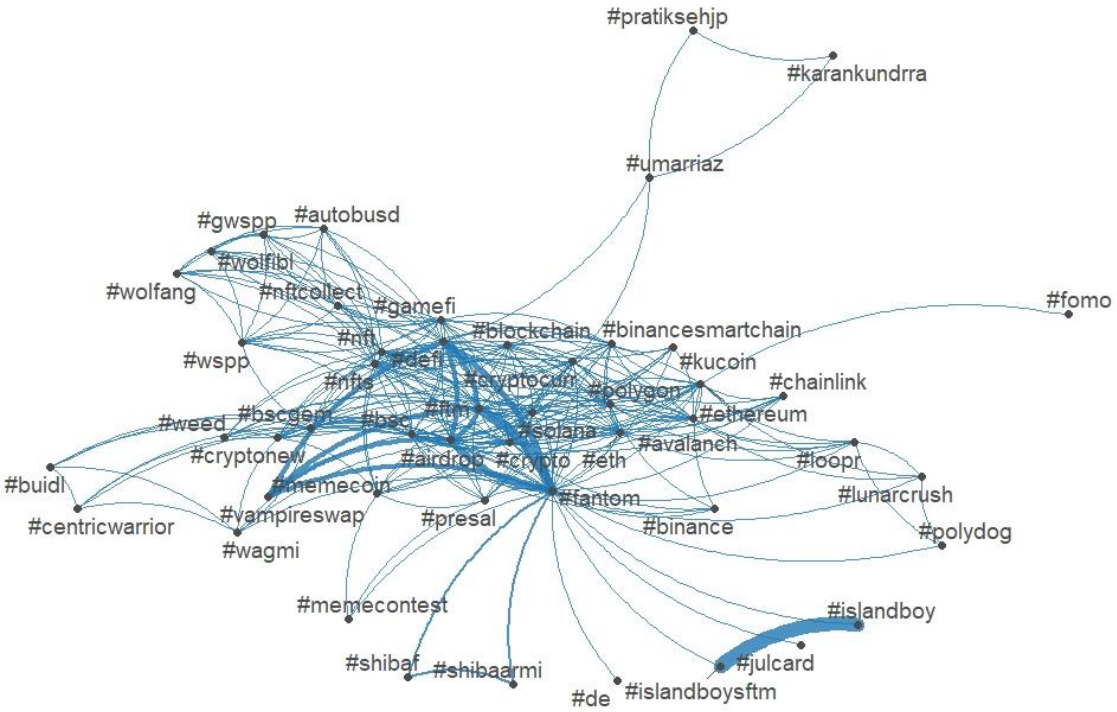
tweet.datetime – the time the tweet was created
text – the text of the tweet
hashtags – use of the # symbol in the text to indicate a subject
currency tags – use of the $ symbol in the text to indicate a subject (specific to currencies)
mentions – use of the @symbol to tag a user in the text

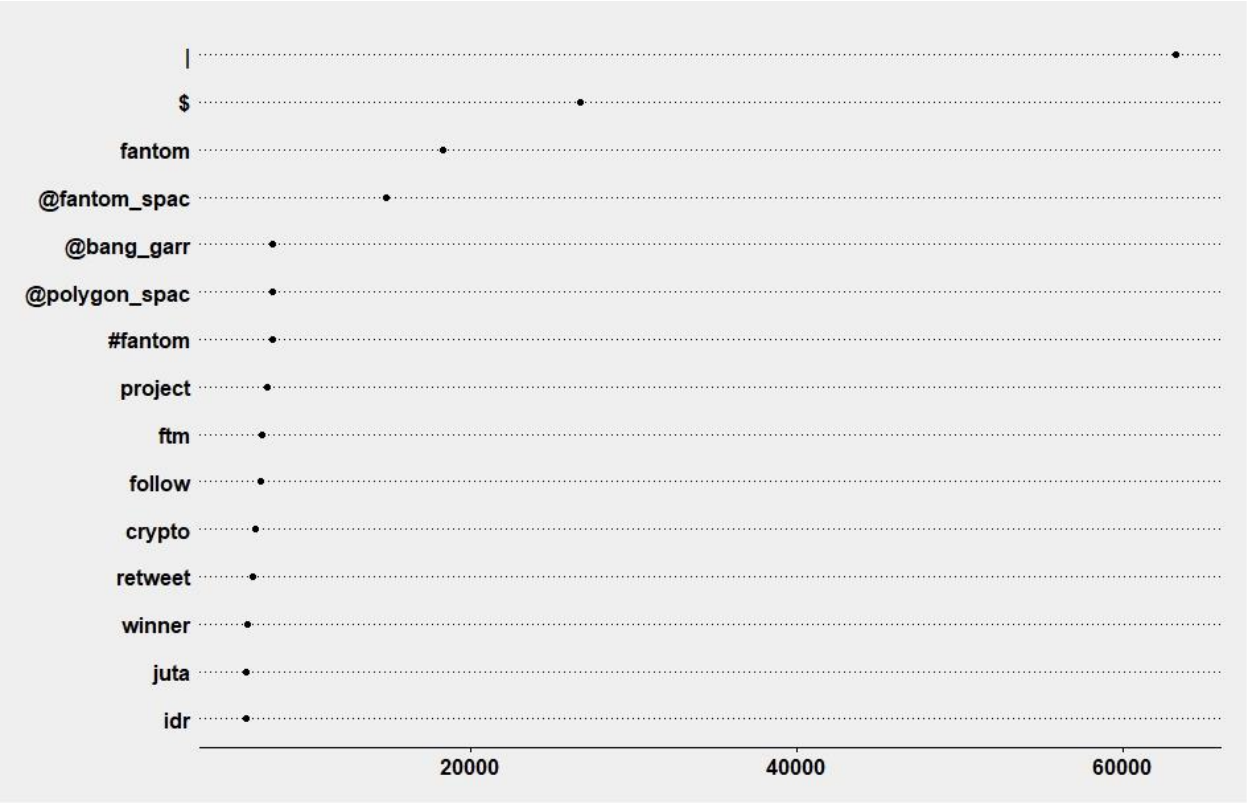
Geckor was used to get price data for the cryptocurrencies evaluated from the website CoinGecko.

The tidyverse suite of packages were used to organize and structure the data in a tibble. The lubridate package was used to clean and process date information. Ggplot2, ggthemes, ggrepel, gridExtra, knitr, kableExtra, formattable, yarr, igraph, ggraph, widyr, and wordclouds were used to convert tabular data into graphical representations. Tidytext, and tm were used to process text data and build natural language processing models. Quanteda was used both for data manipulation, natural language processing, and generating graphical representations of data.

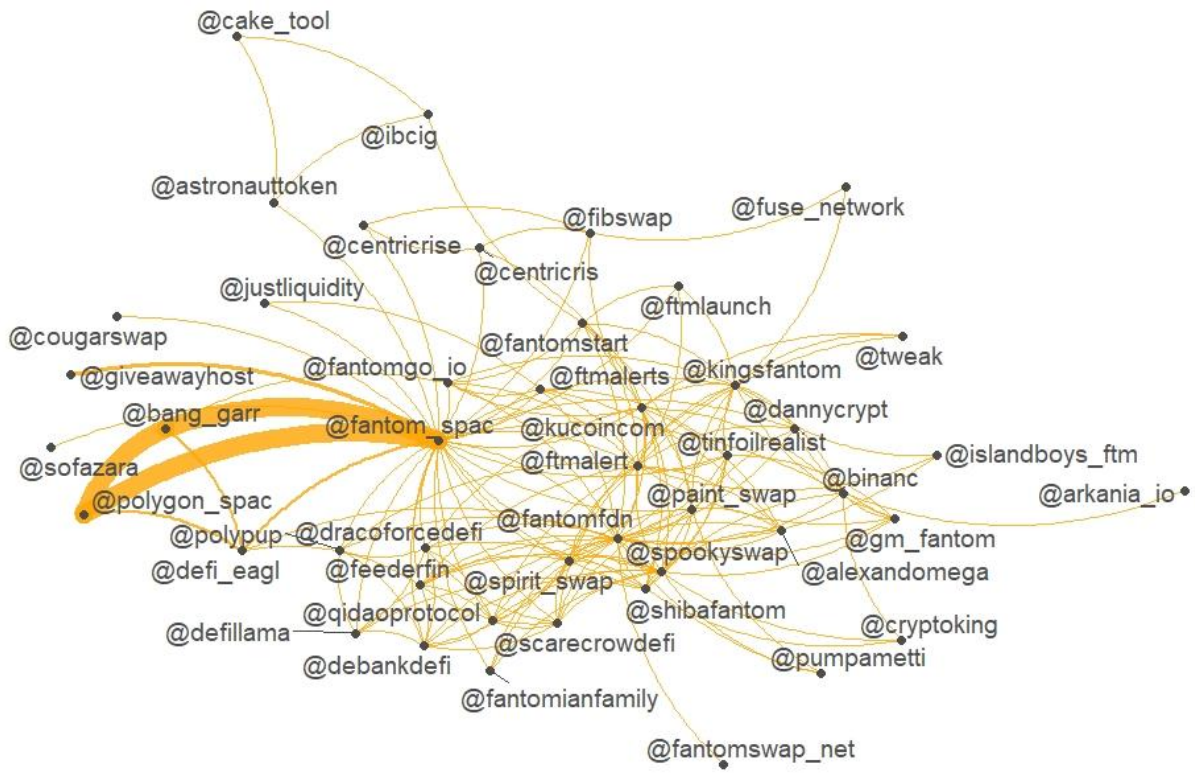
Results

Quanteda was used to identify the most common hashtags, the relationships between hashtags, and the frequency of words across the corpus. These offered some novel insights such as the prevalence of #memecontest and #islandboy. This exploratory data analysis was busier than I'd like and hard to decipher interesting insights.

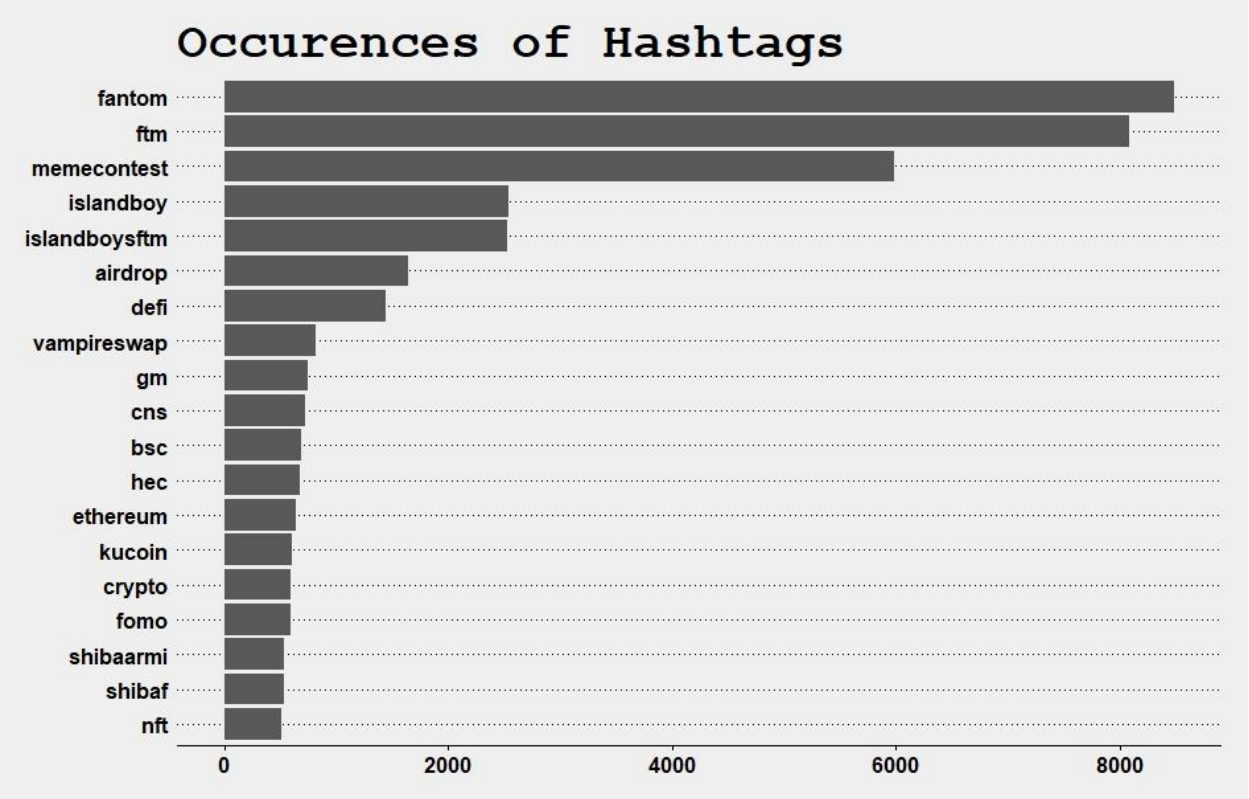




A network focused approach shows some of the most mentioned users and the frequency of their relationships. Future research would seek deeper insights into these relationships. In particular, popular users are often *paid* influencers, so identifying the impact of their messaging may be useful.

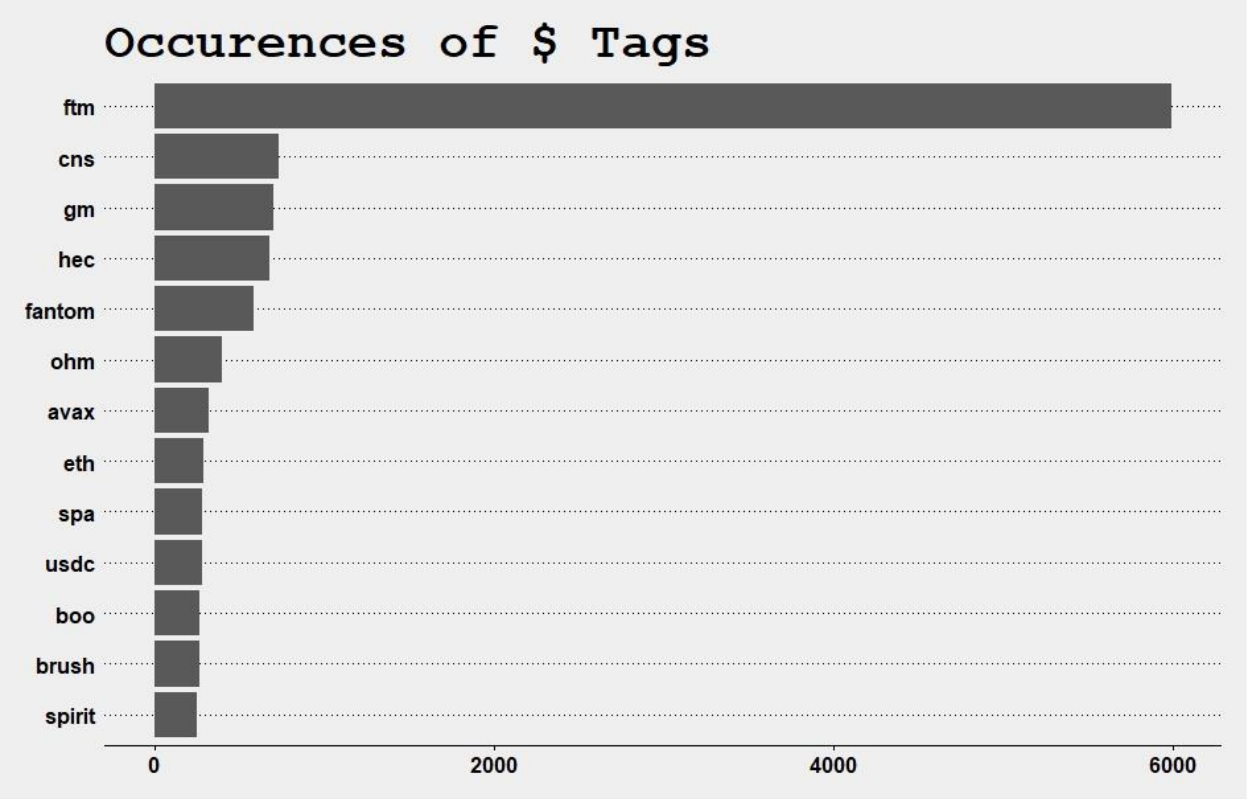


A bag of words analysis of the most common hashtags shows a lot of terms related to the cryptocurrency space. Airdrops are gifted tokens, typically as a reward for users participating in a platform. Defi is the jargon for decentralized finance – users can lend or stake tokens to a protocol and receive interest for doing so or even provide one currency as collateral to borrow another.

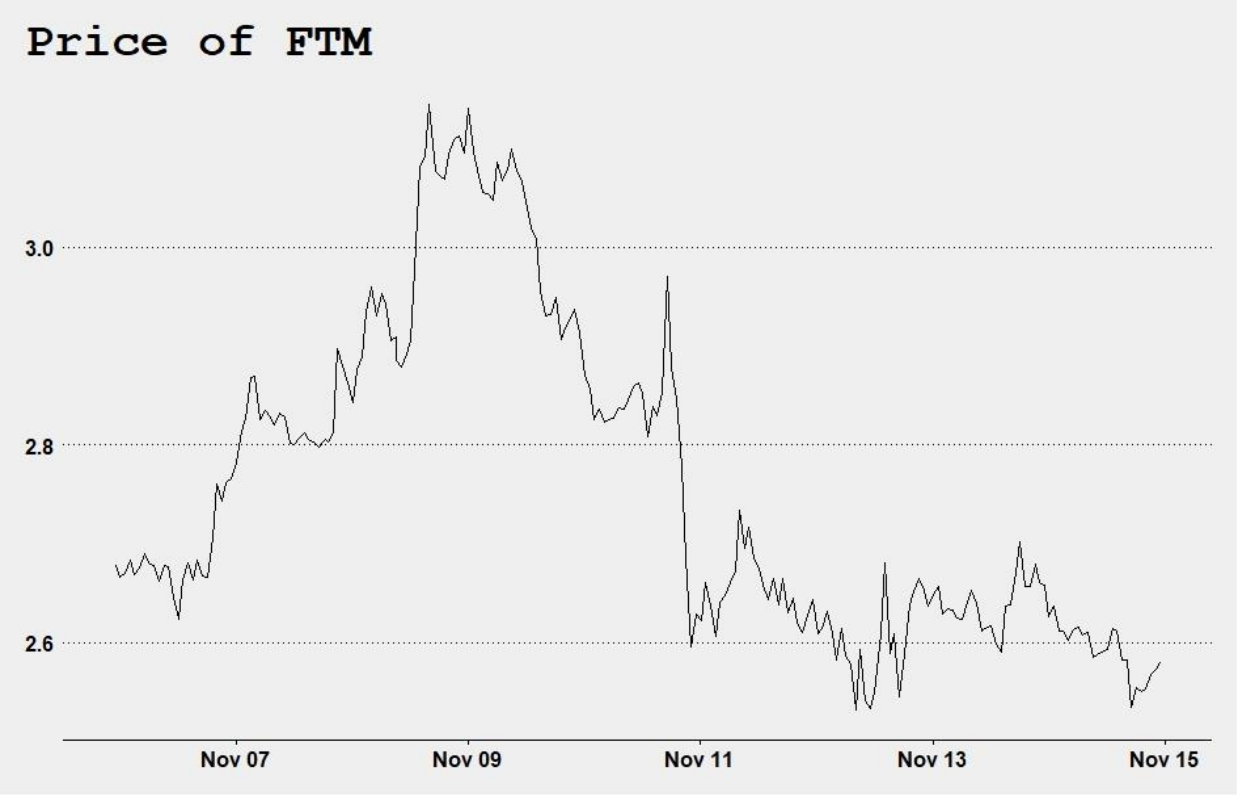


Limiting hashtags to those using the $ flag instead of the # flag shows a very different picture. We would expect ftm as the twitter API query was for fantom. This author is unfamiliar with the token CNS and a CoinGecko search doesn't seem to indicate a link between CNS and Fantom. $GM is a cryptocurrency, but also commonly used as an abbreviation for Good Morning.

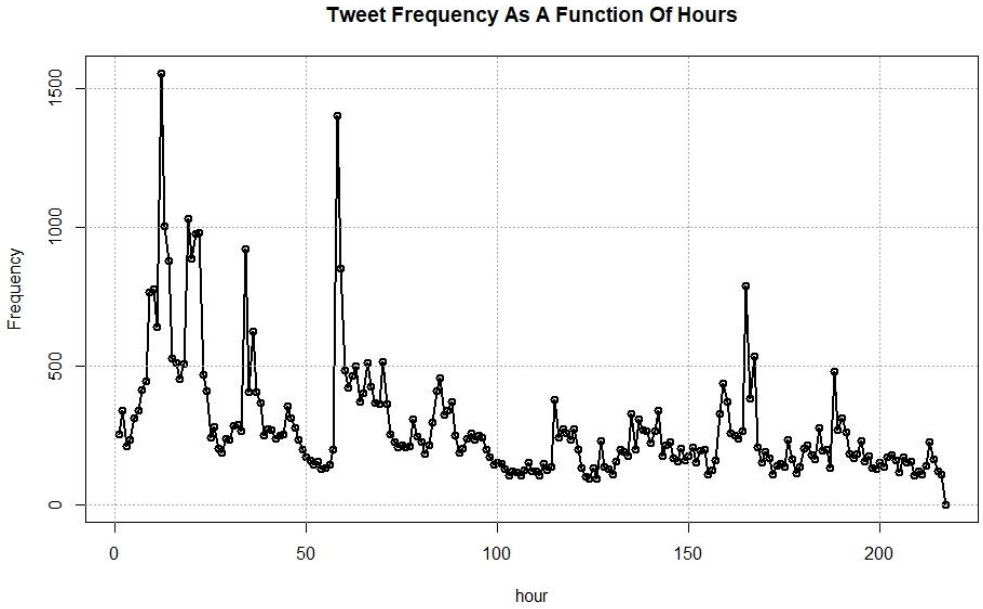
OHM is a popular decentralized reserve currency on the Ethereum blockchain, and during the time of this research, both HEC and SPA were popular "forks" of the OHM protocol on the Fantom ecosystem. BOO and SPIRIT are the tokens for the second and third most popular exchange platforms on the Fantom network (spookyswap and spiritswap). Interesting, GEIST, is the token for the largest exchange, and isn't represented in this list.



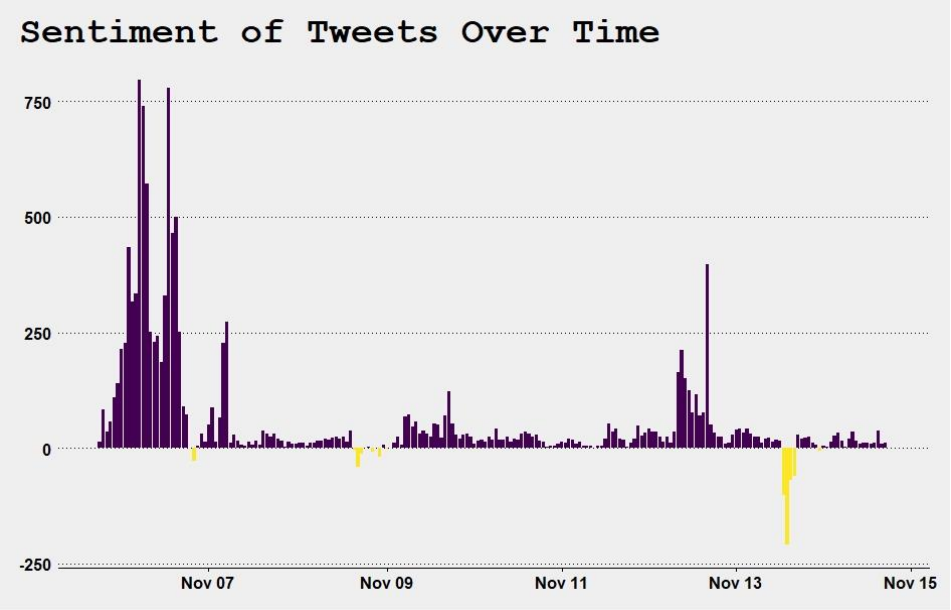
We’re going to consider price as a proxy for adoption of Fantom and/or the protocols in its ecosystem.



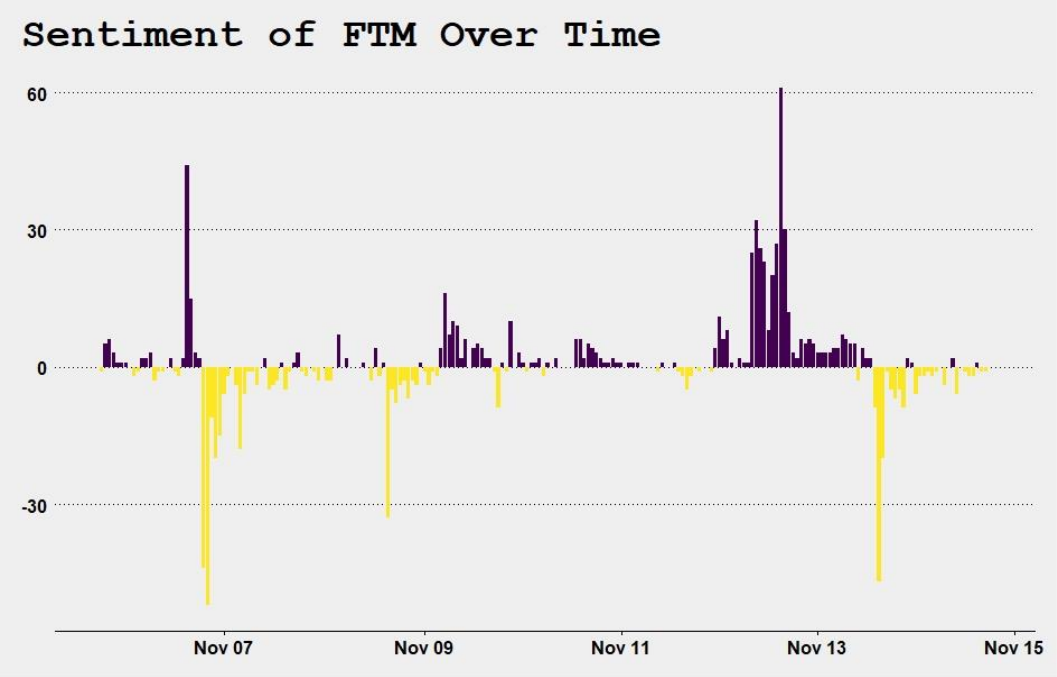
One hypothesis might be that the frequency of tweets might correlate to the price and there does seem to be a correlation between these two charts.



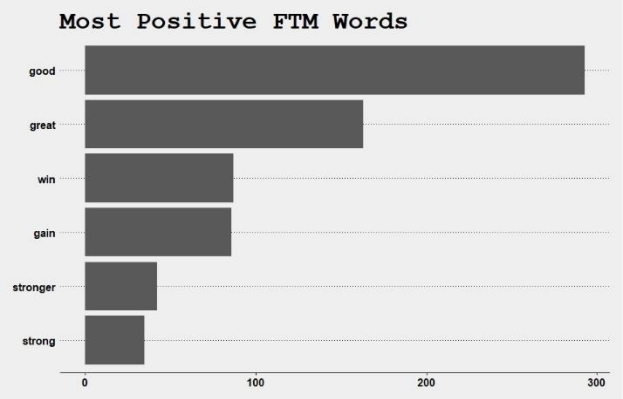
Another consideration might be the sentiment of the tweets over time. For this, the Loughran model was used because it is intended for data related to finances. It seems unlikely that there is a correlation between the two charts. If there is a relationship, it may be delayed (the price rising after the positive sentiment).

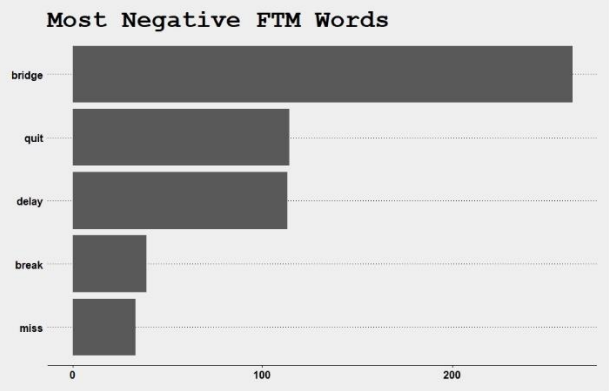


We consider narrowing the sentiment analysis to only tweets using the $ftm hashtag. This paints a very different picture of the sentiment over time. Unfortunately, the possibility of a relationship between sentiment and price seems unlikely.



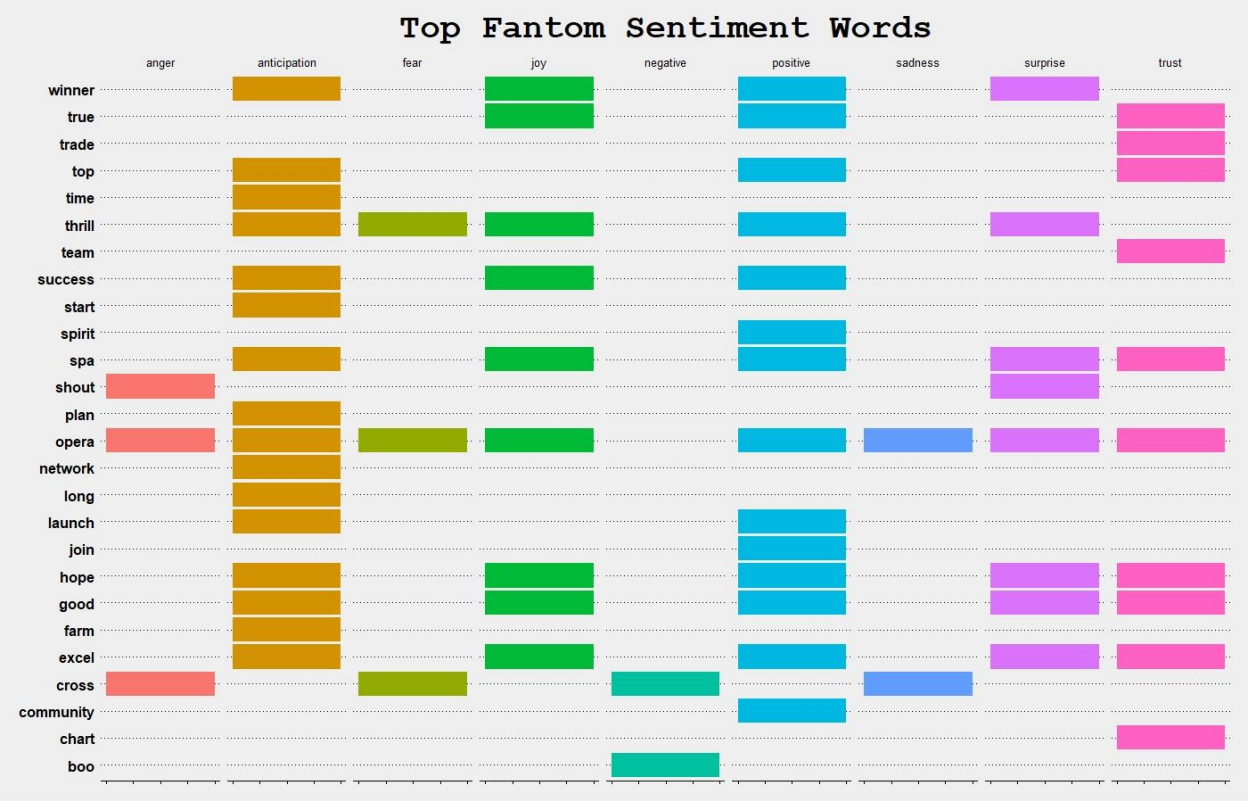
The words associated with positive and negative relationships are evaluated to see if there is anything indicating problems in the model or data. One stand-out is the word “bridge” on the negative relationships. A bridge in cryptocurrency is used to transfer assets from one ecosystem to another and may have a different meaning in other markets.



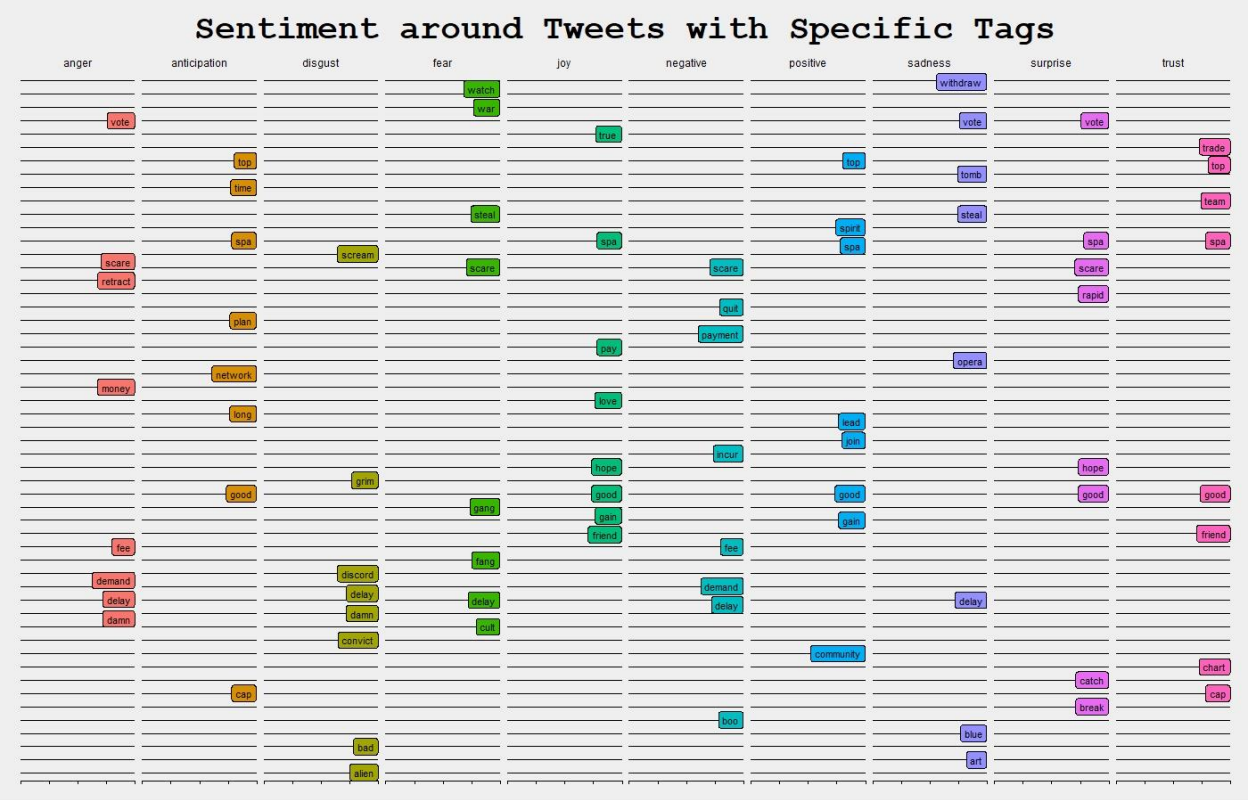


We then look at an NRC model to get a better understanding of the sentiment in the Tweets. Top indicating anticipation, positive, and trust may be appropriate if the user is celebrating something like a new all-time-high. It also could be fear or negative if the user is mentioning buying the top. SPA is the tag for a currency, and should probably have been filtered. Opera is part of the name for the network, “Fantom Opera”.

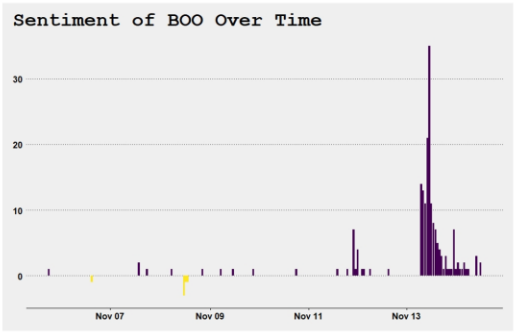
Interestingly, “farming” in cryptocurrency is a name when users provide their tokens as collateral and receive rewards for doing so.

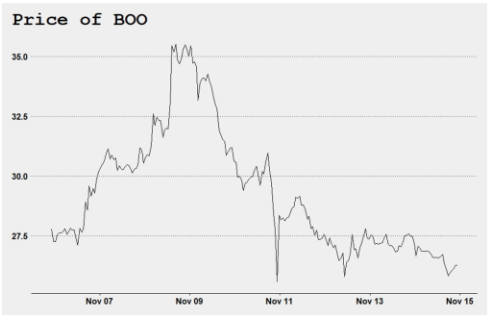


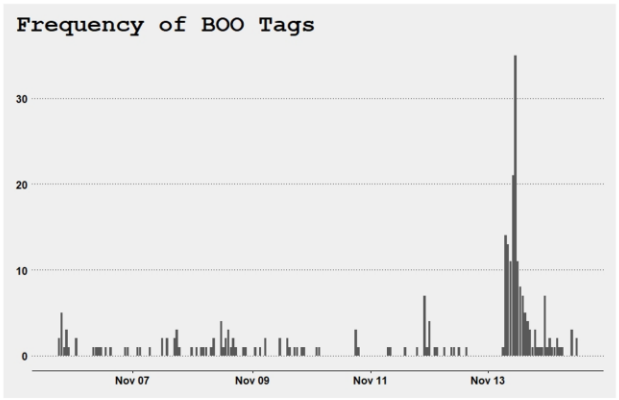
The NRC sentiment was further filtered to just tweets containing the hashtags $ftm, $boo, $spirit, $spa, and $hec as these are all cryptocurrencies in the Fantom ecosystem. A lot of the negative, angry, disgusted, or fearful words are names for platforms within the ecosystem such as grim.finance, scream.finance, discord (a popular communities tool).

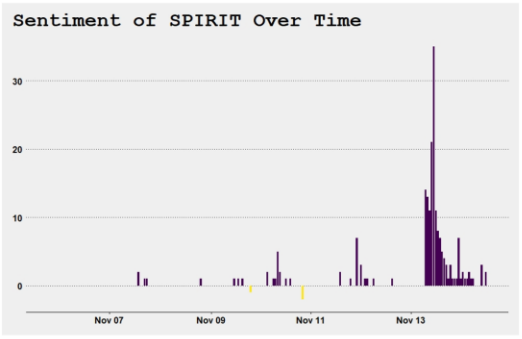


Sentiment, frequency, and price analysis were conducted for SPIRIT, BOO, HEC, and SPA. Bing models were used due to the scarcity of the data. There may be some correlation with HEC, sentiment, and frequency, otherwise, no relationships were indicated.

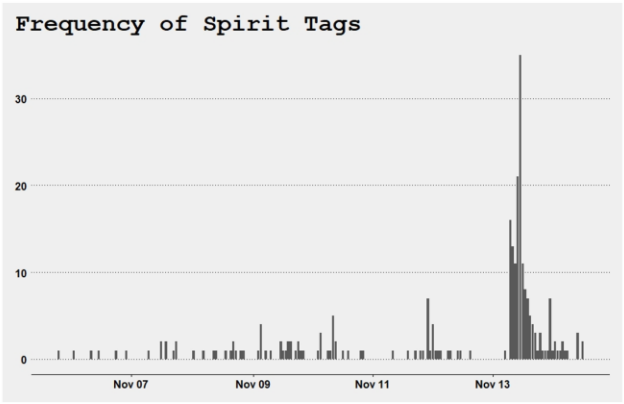

Sentiment of BOO Over Time
30
20
10
0
Nov 07
Nov 09
Nov 11
Nov 13

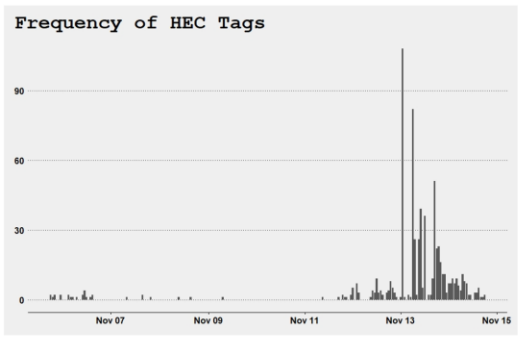

Price of BOO
35.0
32.5
30.0
27.5
Nov 07
Nov 09
Nov 11
Nov 13
Nov 15

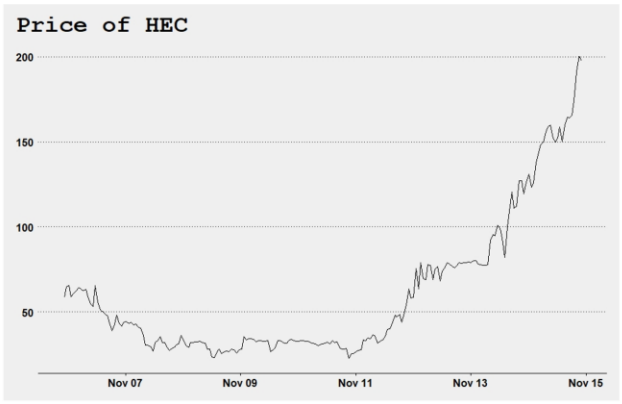

Frequency of BOO Tags
30
20
10
0
Nov 07
Nov 09
Nov 11
Nov 13

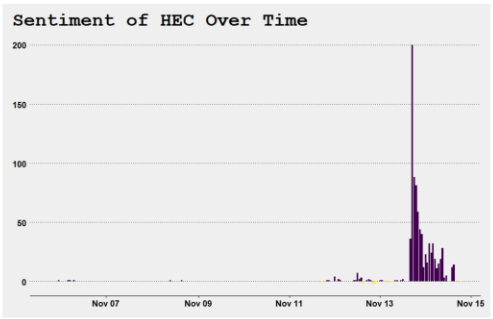

Sentiment of SPIRIT Over Time
30
20
10
0
Nov 07
Nov 09
Nov 11
Nov 13

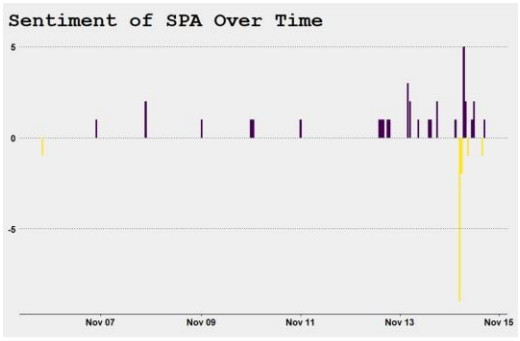

Price of SPIRIT
0.350
0.325
0.300
0.275
0.250
0.225
Nov 07
Nov 09
Nov 11
Nov 13
Nov 15

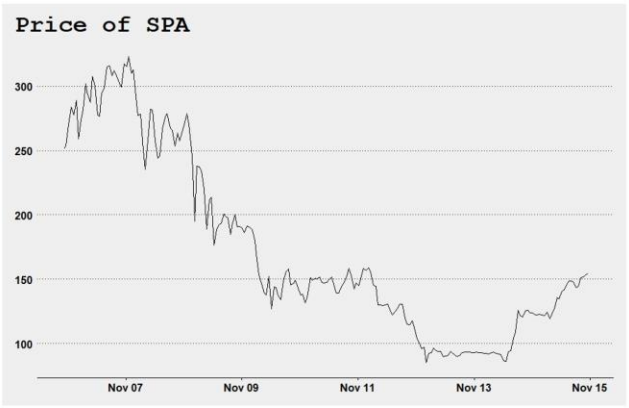

Frequency of Spirit Tags
30
20
10
0
Nov 07
Nov 09
Nov 11
Nov 13

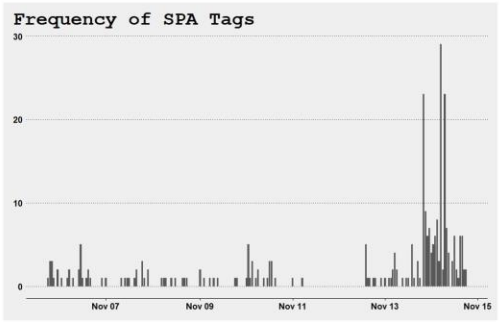






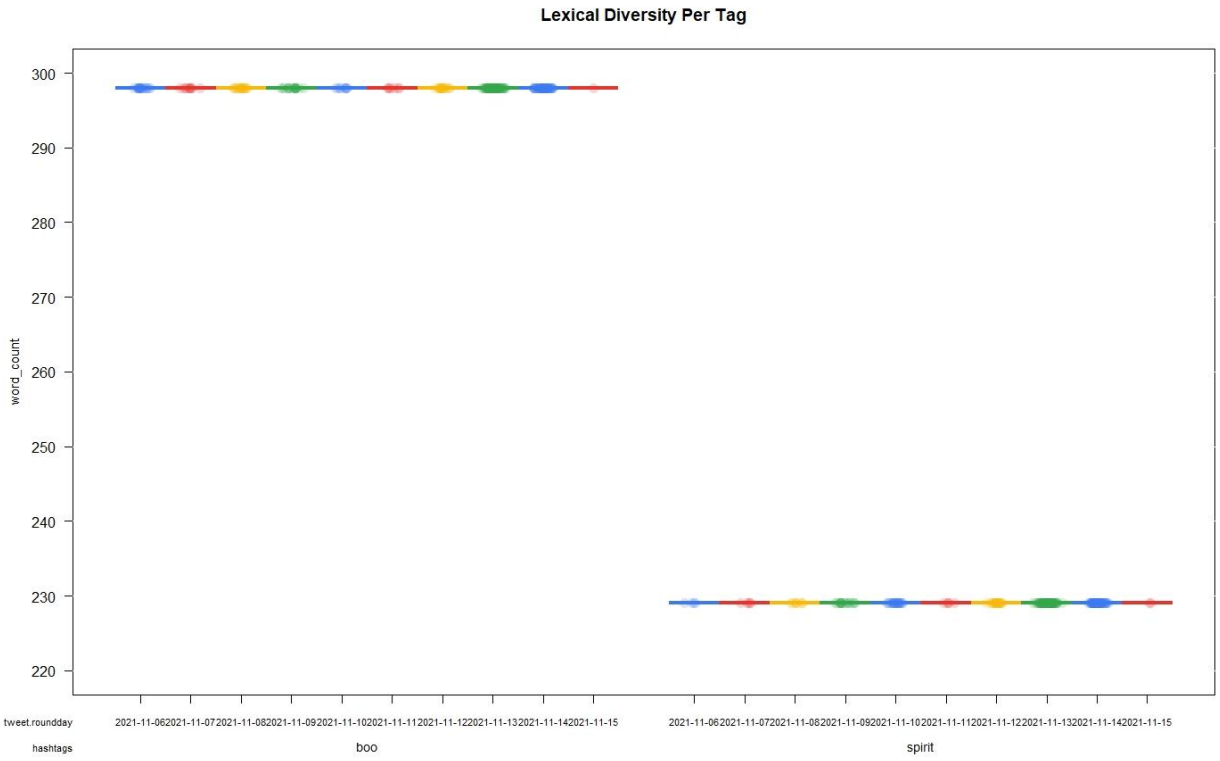






The Lexical diversity of the tweets with $boo and $spirit tags were compared on a day-by-day basis to compare these two platforms. Interestingly, SpookySwap (boo) is the more popular

exchange and the lexical diversity is much higher. Is that product of more tweets or more complex tweets?



Limitations & Next Steps

This project was limited by the use of price per unit as an indicator of diffusion of innovation. The nature of some of these tokens are inflationary while others should be deflationary. Total value invested may have been a better indicator.

This project was also limited by time. 9 days is far too short of a timeline, especially when considering extremely volatile assets like cryptocurrencies. Successful cryptocurrency projects spawn many “forks” of that project – SPA and HEC are both forks of the Olympus DAO project, OHM. Given the open-source nature of development space and the popularity of OHM, hundreds of near carbon copies launched in the month of November.

Furthermore, sentiment analysis is exceedingly difficult in the rapidly evolving jargon space that is Cryptocurrency Twitter and even more complicated when the chosen ecosystem shares so many words with horror genres.

A fun expansion of this project would build, change, and shape the sentiment analysis word lexicon to be more suited to the content, analyze the material over a longer period of time, and seek more network related insights (especially around related users and projects).

Conclusions

Start with the end in mind. This simple idea can really shape how a project grows and expands, especially in this unique space. There seemed to be very little, if any, relationship between tweets and price-per-token; however, the winding path of considering lexical diversity, and sentiment analysis offered more interesting hypotheses and exploratory research.

One popular YouTuber uses the XTZ floor as his basis for the value of a layer 1 ecosystem. His rational, the Tezos (XTZ) ecosystem has nothing on it, so any ecosystem with an actual use case should be, at least, more valuable than Tezos. Currently, Fantom has less market capitalization than Tezos. A lot of the negatively associated terms in this dataset were also names of tokens or platforms in this ecosystem. Could the theme actually be hurting the ecosystem because of the implicit sentiment?

Lexical Diversity offered a really interesting relationship between BOO and SPIRIT. This analysis could be a really interesting way to further explore the popularity and value of other projects, other ecosystems, or even popular users (who may or may not be paid “influencers”).

This was a super fun project. Ending it with more questions has me excited about how much more there is to learn in this space.

References & Packages

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing,
Vienna, Austria. URL https://www.R-project.org/.

Silge J, Robinson D (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." _JOSS_, *1*(3). doi: 10.21105/joss.00037 (URL: https://doi.org/10.21105/joss.00037), <URL: http://dx.doi.org/10.21105/joss.00037>.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Ingo Feinerer and Kurt Hornik (2020). tm: Text Mining Package. R package version 0.7-8. https://CRAN.R-project.org/package=tm

Ingo Feinerer, Kurt Hornik, and David Meyer (2008). Text Mining Infrastructure in R. Journal of Statistical Software 25(5): 1-54. URL: https://www.jstatsoft.org/v25/i05/.

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.

Ian Fellows (2018). wordcloud: Word Clouds. R package version 2.6. https://CRAN.R-project.org/package=wordcloud

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham and Dana Seidel (2020). scales: Scale Functions for Visualization. R package version 1.1.1. https://CRAN.R-project.org/package=scales

Jeffrey B. Arnold (2021). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.4. https://CRAN.R-project.org/package=ggthemes

Achim Zeileis and Gabor Grothendieck (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27.

doi:10.18637/jss.v014.i06

Hadley Wickham (2021). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. https://CRAN.R-project.org/package=forcats

Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. https://igraph.org

Thomas Lin Pedersen (2021). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 2.0.5. https://CRAN.R-project.org/package=ggraph

David Robinson (2021). widyr: Widen, Process, then Re-Tidy Data. R package version 0.1.4. https://CRAN.R-project.org/package=widyr

Grün B, Hornik K (2011). "topicmodels: An R Package for Fitting Topic Models." _Journal of Statistical Software_, *40*(13), 1-30. doi: 10.18637/jss.v040.i13 (URL: https://doi.org/10.18637/jss.v040.i13).

Murzintcev Nikita (2020). ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. R package version 1.0.2. https://CRAN.R-project.org/package=ldatuning

Bob Rudis, Niels Lohmann, Deepak Bandyopadhyay and Lutz Kettner (2019). ndjson: Wicked-Fast Streaming 'JSON' ('ndjson') Reader. R package version 0.8.0. https://CRAN.R-project.org/package=ndjson

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018). "quanteda: An R package for the quantitative analysis of textual data." _Journal of Open Source Software_, *3*(30), 774. doi: 10.21105/joss.00774 (URL: https://doi.org/10.21105/joss.00774), <URL: https://quanteda.io>.

Sergey Mastitsky (2021). geckor: R Client for the

'CoinGecko' API. R package version 0.2.0.
https://CRAN.R-project.org/package=geckor

Kamil Slowikowski (2021). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.1.
https://CRAN.R-project.org/package=ggrepel

Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3.
https://CRAN.R-project.org/package=gridExtra

Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.35.

Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595

Hao Zhu (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4.
https://CRAN.R-project.org/package=kableExtra

Kun Ren and Kenton Russell (2021). formattable: Create 'Formattable' Data Structures. R package version 0.2.1.
https://CRAN.R-project.org/package=formattable

Nathaniel Phillips (2017). yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". R package version 0.1.5. https://CRAN.R-project.org/package=yarrr