ISSS606 – Social Analytics and Applications
Project Proposal

NFT NETWORK ANALYSIS

GROUP 5:

ASM
CS
AG
HD
PRA
ZX

Introduction

Since its inception in 2014, Non-Fungible Tokens (NFTs) were privy only to a few elite crypto aficionados. However, Beeple's $69 million deal in March 2021 placed NFTs on the map for the public at large. NFTs are the harmonious blend of art and cryptocurrencies, in the form of art, music, videos, images, etc. Their underlying blockchain technology is like Bitcoin's, but each NFT has its own unique digital signature which renders an exchange with another NFT impossible. The NFT network is exclusive with high barrier to entry and the commodity value is largely a community driven endeavor. With NFTs growing, currently the general public's opinion is polarized with many wondering if it is the key to the metaverse or just overpriced JPEGs.

To begin with, we researched recent studies that analyzed how social media, a popular medium for exchanging ideas for these assets, can impact the NFTs' valuation (Kapoor et al., 2022). Thus, for our study, we aimed to uncover insights on the overall sentiment of NFT market and deep dive into their network or communities between artists and the collectors.

Objective

The objective of this project is to perform NFT network analysis. To do so, the following tasks were identified:

1. **Perform sentiment analysis on the overall NFT network**: Gain understanding on the general sentiment towards NFT's by the NFT network identified.
2. **Perform community detection**: The NFT network is community driven. The purpose of the community detection is to effectively determine the different NFT communities.
   i. Perform sentiment analysis: Gain understanding on the sentiment towards NFTs by the different communities detected.
   ii. Perform Topic modelling: Gain insight into how different communities give importance to certain topics and how it affects the NFT market.

Dataset

To perform the objectives mentioned in this project, we decided to use Twitter posts related to NFT's. We used two major sources which has been summarized in the table below:

Type Of Data	Source	Data Description	Purpose
Social media data	Verified NFT Tweets from Kaggle (Ada-Nai, 2022)	No. of tweets: 94897 From: 01/01/2021 To: 31/12/2021 Contains 84,416 English tweets	The purpose of these tweets is to perform sentiment analysis, community detection and topic modelling.
	Twitter API (with tweepy python package)	q = "#nft -filter:retweets filter:verified since:" Here, we considered tweets from users verified since 2021-01-21	

Commented [ZX1]: I think we use more data from twitter from mid May? to June? need to check the data

In the above table we can see the term NFT Tweets. This refers to tweets with the hashtag nft (#nft).

The Kaggle dataset contains NFT tweets from verified users only. Similarly, while extracting the NFT tweets using the twitter API, we ensured to only extract tweets from verified users as shown in the table above.

Extracting tweets of verified users allows us to avoid bots. We also didn't consider retweets in this project. Retweets are just repetitions of the same tweet. By not including retweets, we can make sure the sentiment of the tweets are not inflated.

Methodology & Analysis

Background: Prior to 21 February 2021, NFTs were privy to smaller communities familiar with the blockchain and the small artist network that leveraged NFTs to reserve rights to their art. However, it was Beeple's 'The first 5000 days' that put NFTs on the global map - the inordinately large sale worth USD 69 million intrigued people world over. The bulk of NFTs run on Ethereum blockchain as it enables NFTs to store more information to function distinctly from other crypto currencies. Thus, one can say that the price of ETH and NFTs are inextricably linked. Since market sentiments are largely responsible for the volatility in stocks/crypto, it would be worth checking if there is any link between the ETH price movements to NFT sentiments in the overall network as well as looking into the co-movement of ETH and NFT sentiments within each community.

Ethereum Price movement: Since 21 February 2021, there has been an increase in the price of Ethereum. The first peak was observed on 10 May 2021 where the price of ether crossed the USD 4,000 threshold for the first time, making it the second largest cryptocurrency – an event attributed to the popularization of NFTs (Browne, 2021). Although, this elation was rather short-lived as the price experienced a sharp decline in the following June due to market turbulence and high volatility of ETH. However, the market remained bearish as the decline leveraged traders to buy expensive ETH at a lower price (Brockman, 2021). Subsequently, ETH recovered, and its price soared to an all-time high in the following November. ETH's support for Metaverse, adoption surge of NFTs, growth of decentralized finance, and increase in ETH developers are potential factors that drove up the prices (Upadhyay & Upadhyay, 2021). The steady decline post November could be attributed to covid peak and other unknown factors.

Sentiment Analysis of overall network

Commented [ZX2]: i think it is better to combine them toghter as a methodology part, and in it, we introduce our challenges and our final choice

Commented [PA3R2]: Done

Challenges with using Vader: To perform sentiment analysis, we explored the option of using vader, a pretrained sentiment model from the nltk package. However, we found some challenges:

1. Misinterpretation of sarcasm or irony

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
analyzer = SentimentIntensityAnalyzer()
sentence = "Covid cases are increasing fast!"
analyzer.polarity_scores(sentence)

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Figure 1 - Vader - Example 1

The above sentence should have been classified as negative. However, it is incorrectly classified as neutral.

2. Use of emojis affects the sentiment result

```
sentence = "NFT to the moon 🚀🚀"
analyzer.polarity_scores(sentence)

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Figure 2 - Vader - Example 2

The above sentence should have been classified as negative. However, it is incorrectly classified as neutral.

Moreover, the language and grammar used in tweets don't typically represent the language rules and conformations of the English language. This might also arise from the fact that tweets are at most only 280 characters long. Thus, Vader may not always give satisfactory results when applied to tweets. This thus leads us to exploring alternate options to Vader.

Exploring alternative solutions:

TimeLMs Sentiment Analysis: We explored the option of using TimeLMs. It is a pretrained "RoBERTa-based model trained with about 124M tweets" (Cardiffnlp, n.d.). The model yields a probability score for the tweet to be positive, neutral, negative where the highest score will be the chosen sentiment. These pretrained models take context and sentence structures into consideration (Loureiro et al., 2020). Since this model is trained with tweets, its performance expectation is higher. Thus, we chose this to be our sentiment analysis model.

Result and analysis

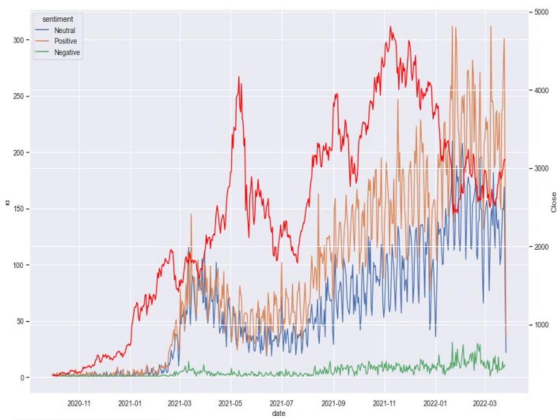

Figure 3 – Sentiment Analysis of overall NFT network

As can be seen in the figure above, we can note that the reaction towards NFTs picked up in at Quarter 2 of 2021. If we focus our attention at Quarter 1 of 2022, we see that while the negative comments are low in proportion, it is more evident.
Also, as mentioned earlier, we wanted to check if there is any link between the ETH price movements to the sentiments in this network. In this case, we can observe an interesting insight that the sentiments are not tied to the ETH price.

In summary, we can note that the twitter sentiments are mainly positive in the observed period.

Error Analysis:
As TimeLMS is not trained specifically with NFT related tweet, it won't be a perfect model. Below table summarized the errors based on 100 samples for each sentiment category.

Sentiment	Accuracy	Topics	Errors
Positive	**96%**	• 'drop' is recognized as positive term • Promoting new tokens • Value • Decentralization	• Mainly neutral, factual comments
Negative	**86%**	• Scams & spams • Errors in transaction/minting • Gas fees • Skepticism of technology	• "sold out" • "and you say NFT has no use case" • Name of token "Super Virus" • "Don't understand web3"
Neutral	**84%**	• Factual news/statements • New platform features • Technical questions ("moving tokens") • Content with only links	• Promotion of NFTs • New tokens drop

From the above, we can ascertain that the accuracy is decent.

Community Detection

For a novice person wishing to enter the NFT world, one thing is a must – understanding the deep-seated communities that make the wheels turn. They are responsible for generating a raging milieu of creators and collectors, who propel the NFT markets to their current glory. For our analysis, we performed community detection, then sentiment analysis and topic modelling for the top communities to give a holistic view on each community.

Methodology for detecting the top 8 communities

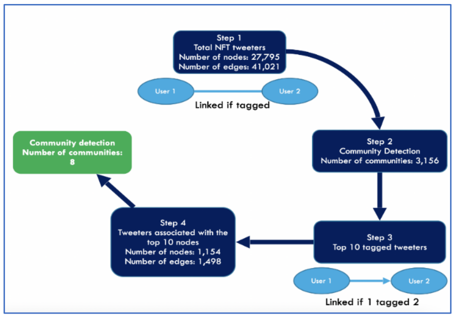

Figure 44 – Methodology for top 8 community detection

First, we built a community network. For this, we identified tweeter users with #nft tweets and their corresponding linked users. Next, we used modularity-based agglomerative algorithm to detect over 3,000 communities. However, given that such a large network makes it challenging to analyze and interpret insights, we built a directed network. For a directed network, we consider a user as an influence if other users tag him/her in their NFT tweets. This helped us the size and the number of communities. Using PageRank algorithm, we selected top 10 influencers and their corresponding communities for further analysis. This helped obtain relatively smaller networks and the nodes/users associated with these top 10 influencers. Ultimately, we were able to detect eight communities based on the directed network of the top 10 influencers.

Result and analysis

Community 0 – Rariblecom: It is a minting as well as a trading platform, founded in 2020. Most recent of the lot, it has exacted considerable influence in the NFT world. It's unique selling point is lies the addition of hidden messages and resolution files in their art. It has 480 thousand followers on twitter.

Community 1 – OpenSea: It currently dabbles in a wide genre of NFTs such as art, collectibles, music, photography, sports, etc. with over 20 million NFTs to work with. It currently boasts of 1.8 million followers on twitter and is a must know platform for any novice player. It is an interactive marketplace with statistics such as top NFT trends, and collections by categories and chains.

Community 2 – Niftygateway: It is an exclusive NFT marketplace for music and art. It has curated drops from verified collectors. The NFTs sold here are called Nifties! It currently boasts of 250,000 followers on twitter.

Community 3 – Garyvee: Gary Vaynerchuk is a touted social media influencer who has 3.1 million followers on twitter. While he is known for his motivational speaking and business acumen, he is also known to carry considerable significance in the NFT market. He has his own line of NFTs called VeeFriends which in its first iteration has garnered around 49,000 ETH (approximately $160 million).

Rarible, Christiesinc, Metamask, Coin artists have significant impact in the NFT marketplace.

Top 8 Communities

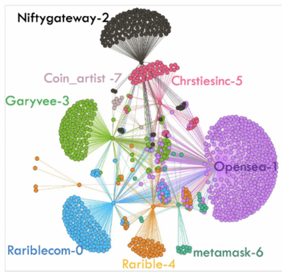

Figure 55 - Top 8 communities

We performed sentiment analysis and topic modelling for the communities detected to get a pulse of how movements in the NFT markets are being perceived by different communities.

Sentiment Analysis on the Communities detected

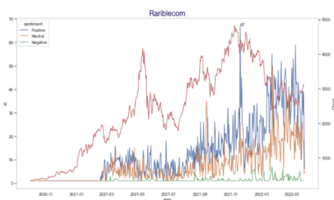

Figure 6 6 – Rariblecom

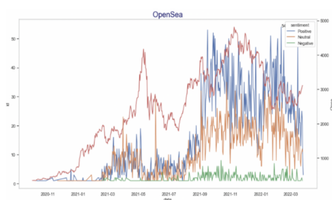

Figure 7 - OpenSea

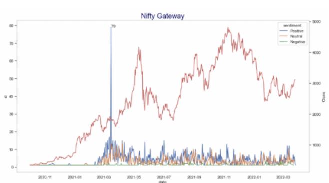

Figure 8 - Nifty Gateway

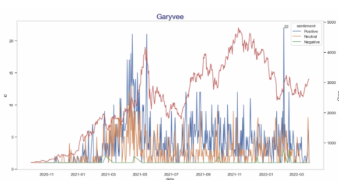

Figure 9 – GaryVee

Note: The red line is the price of Ethereum, blue line is the neutral sentiment, orange is the negative sentiment and green is the positive sentiment.

Community Analysis

Before NFT market gained momentum, it relied on Ethereum to lend it value. However, as it has been gaining popularity, it appears to be branching away. NFTs react favorably to unfavorable macroeconomic factors because of the role it essays in the metaverse and 'play-to-earn' narratives it supports, decline in the popularity of bitcoin and shifting focus and faith to Ethereum (Thompson, 2022).

This phenomenon is echoed by the eight communities for which sentiment analysis was conducted. For instance, OpenSea has been the go-to marketplace since its advent and its positive sentiments were consistent with ETH throughout the period except, they were at an all-time high during September 2021 to March 2022, despite the rise and fall of Ethereum in that period. MetaMask observed similar behavior. In contrast, Rarible and Rariblecom observed a peak positive sentiment after the slight decline in Ethereum prices in November (after ETH crossed USD 4,500) showing some correlation between ETH and NFTs. This can be attributed to the platform's users desire to leverage smart buying during the price 'slump', right after the rise. Nifty gateway users did something similar in March 2021. Garyvee, on the other hand, followed a logical correlation between Ethereum price and NFTs in April 2020 and February 2022 but still had positive sentiments during ETH's first peak in June.

Casting Ethereum price aside, the sentiment for most communities has been positive and neutral implying that the influencers have managed to restore the faith in NFTs despite its volatility. The negative sentiments have been characterized as infrequent and with low variance.

Topic Modelling

Topic Modelling is an unsupervised machine learning model that clusters words for a set of documents. We performed topic modelling to determine the pulse of conversation within each community. For our analysis, we used the BERTopic technique.

BERTopic Dynamic Topic Modeling: BERTopic utilized Sentence-BERT word embedding model to perform topic modeling, it has the flexibility on using different pretrained model and clustering techniques (Grootendorst, 2022). Before ingesting the data, a simple preprocessing is performed (e.g., remove mentions, symbols). This not only improves the relevancy of results, but also the processing speed. The first step is to embed our tweets into word vectors to cluster semantically similar tweets. Dimension reduction, UMAP, is applied to reduce the dimensionality of the embedded documents before applying HDBSCAN clustering. A unique element of HDBSCAN includes outliers not being clustered – eventually excluding outliers from interpretation. Finally, a class based TFIDF representations are generated and improved with maximal marginal relevance coherence to minimize overlaps of words.

However, as the model requires embedding, it will be computationally more expensive than LDA/NMF methods. Furthermore, as the length of tweets is short, we didn't not consider using LDA for topic modelling.

Result and analysis

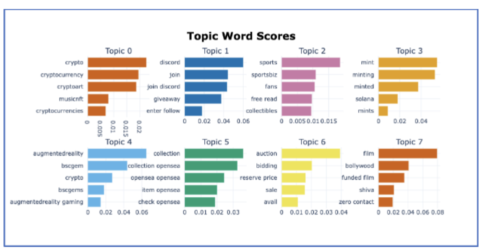

Figure 10 – Rariblecom (Topic Word Scores)

Community	Common Topics	Unique Topics
Rariblecom	Crypto, minting, auction	Sports, augmented reality, Bollywood films
Opensea	Minting, crypto, discord, auction, artist	Cronos, diamonds
Niftygateway	Minting, crypto, discord, auction, drop, artist	shopping, space
Garyvee	Minting, crypto, discord, auction, drop, artist	Halloween, clubhouse, metaverse
Rarible	Minting, crypto, discord, auction, drop, artist	pokeys, opensea, women, music
Chrtiesinc	Crypto, auction	games, newslisting, rendering network
Metamask	Minting, artist	Coolcat, world, fashion, twitter

With the help of this analysis, an NFT enthusiast can select which community they identify with the most. This furthers impassioned discussion within communities, builds culture and helps creators and collectors alike.

Conclusion & Future Works

As can be seen in our project, both sentiment analysis and topic modeling leverage on pretrained models. This allows adoption of newly developed advanced language models whenever available. Such an approach offered us holistic insights about the impact of Ethereum prices on the NFTs, analyzing sentiment of the overall NFT network, detecting top communities, gauging the topics of discussion among different communities as well as get a pulse of sentiments within these communities. We then layered our technical results with secondary research to generate deeper insights. Such an approach, with amplified objectives and real-time data, can be used withing industries where discussion are community driven with substantial impact on the overall market.

From limitations and future work perspective, we believe that the sentiment analysis and the community detection that we performed for the purpose of this project are not trained specifically for NFT tweets. Expanding the universe of hashtags can help draw more tweets for the purpose of analysis. Additionally, twitter individually does not represent the sentiment of the entire NFT market. Including texts and node network from other platforms where discussions related to NFTs are dominant can help enhance the quality of insights and offer a well-rounded view of the marketplace.

Team Contribution

Milestone	Members
Ideation	Prachi, Aishwarya, Architha
Conceptualization and planning	Prachi, Aishwarya, Architha, Huo Da, XinXin, Shaoyao
Data Crawling	Prachi, Aishwarya, Architha, Huo Da, XinXin, Shaoyao
Coding and execution	Huo Da, Shaoyao, XinXin, Aishwarya
Report Writing and Presentation Development	Prachi, Aishwarya, Architha, Huo Da, XinXin, Shaoyao

References

Ada-Nai. (2022, April 11). *Verified NFT tweets*. Kaggle. Retrieved June 23, 2022, from https://www.kaggle.com/datasets/adanai/verified-nft-tweets?select=verified_nft_tweets.csv

Brockman, K. (2021, June 24). Ethereum is crashing: Is now the time to invest? The Motley Fool. Retrieved July 5, 2022, from https://www.fool.com/investing/2021/06/24/ethereum-is-crashing-is-now-the-time-to-invest/

Browne, R. (2021, May 10). Ethereum, the world's second-largest cryptocurrency, soars above $4,000 for the first time. CNBC. Retrieved July 5, 2022, from https://www.cnbc.com/2021/05/10/ethereum-eth-price-soars-above-4000-for-the-first-time.html

Cardiffnlp/twitter-roberta-base-sentiment-latest · hugging face. cardiffnlp/twitter-roberta-base-sentiment-latest · Hugging Face. (n.d.). Retrieved June 23, 2022, from https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *Computation and Language*. https://arxiv.org/abs/2203.05794

Loureiro, D., Barbieri, F., Neves, L., Anke, L. E. & Camacho-Collados, J. (2022). TimeLMs: Diachronic language models from Twitter. https://doi.org/10.48550/arXiv.2202.03829

Thompson, N. (2022, March 7). It's complicated: The relationship between crypto and nfts. CoinDesk Latest Headlines RSS. Retrieved July 5, 2022, from https://www.coindesk.com/layer2/2022/03/07/its-complicated-the-relationship-between-crypto-and-nfts/

Upadhyay, R., & Upadhyay, R. (2021, November 8). Ether price analysis 8 Nov: Can bulls ascend $5,000? Ether price analysis 8 Nov: can bulls ascend $5,000? | Currency.com. Retrieved July 5, 2022, from https://currency.com/ether-price-analysis-november-8-2021