Deciphering Crypto Twitter

Inwon Kang
Rensselaer Polytechnic Institute
Troy, New York, USA
kangi@rpi.edu

Maruf Ahmed Mridul
Rensselaer Polytechnic Institute
Troy, New York, USA
mridum@rpi.edu

Abraham Sanders
Rensselaer Polytechnic Institute
Troy, New York, USA
sandea5@rpi.edu

Yao Ma
Rensselaer Polytechnic Institute
Troy, New York, USA
may13@rpi.edu

Thilanka Munasinghe
Rensselaer Polytechnic Institute
Troy, New York, USA
munast@rpi.edu

Aparna Gupta
Rensselaer Polytechnic Institute
Troy, New York, USA
guptaa@rpi.edu

Oshani Seneviratne
Rensselaer Polytechnic Institute
Troy, New York, USA
senevo@rpi.edu

ABSTRACT

Cryptocurrency is a fast-moving space, with a continuous influx of new projects every year. However, an increasing number of incidents in the space, such as hacks and security breaches, threaten the growth of the community and the development of technology. This dynamic and often tumultuous landscape is vividly mirrored and shaped by discussions within "Crypto Twitter," a key digital arena where investors, enthusiasts, and skeptics converge, revealing real-time sentiments and trends through social media interactions. We present our analysis on a Twitter dataset collected during a formative period of the cryptocurrency landscape. We collected 40 million tweets using keywords related to cryptocurrency and performed a nuanced analysis that involved grouping the tweets by semantic similarity and constructing a tweet and user network. We used sentence-level embeddings and autoencoders to create K-means clusters of tweets. We identified six groups of tweets and their topics to examine different cryptocurrency-related interests and the change in sentiment over time. For example, we identified different groups of tweets demonstrating coordinated behavior in the market or expressing distrust in centralized cryptocurrency exchanges. Moreover, we discovered sentiment indicators that point to real-life incidents in the crypto world, such as the FTX incident of November 2022. We also constructed and analyzed different networks of tweets and users in our dataset by considering the reply and quote relationships and analyzed the largest components of each network. Our networks reveal a structure of bot activity in Crypto Twitter and suggest that they can be detected and handled using a network-based approach. Our work sheds light on the potential of social media signals to detect and understand crypto events, benefiting investors, regulators, and curious observers alike, as well as the potential for bot detection in Crypto Twitter using a network-based approach.

CCS CONCEPTS

• **Human-centered computing** → **Social network analysis**; • **Networks** → **Social media networks**; • **Information systems** → *Content analysis and feature selection*; *Sentiment analysis.*

KEYWORDS

Cryptocurrency, Blockchain, Twitter, NLP, Social Networks

ACM Reference Format:
Inwon Kang, Maruf Ahmed Mridul, Abraham Sanders, Yao Ma, Thilanka Munasinghe, Aparna Gupta, and Oshani Seneviratne. 2024. Deciphering Crypto Twitter. In *ACM Web Science Conference (WEBSCI '24), May 21–24, 2024, Stuttgart, Germany*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3614419.3644026

1 INTRODUCTION

The dynamic and often unpredictable nature of the cryptocurrency market, where rapid innovation intersects with complex security challenges, has rendered it a unique domain for exploring the interplay between technological developments and public sentiment, particularly as mirrored in the vast digital landscape of social media. Following the release of Ethereum in 2015, the crypto world has gained access to ever-increasing programmability in the blockchain, leading to the *Initial Coin Offerings* (ICO) craze of 2017 [11, 15]. While that excitement has toned down, a steady stream of new projects in the cryptocurrency world are constantly "advertised" and discussed on social media.

However, recent years have seen an increasing number of attacks on blockchain projects [5], much of those stemming from social engineering attacks that originate on social media. CNBC reports that nearly $4 billion were stolen in just 2022 [9], and the number of lawsuits related to these losses continuously increased, reaching up to 20 suits filed annually by 2023 [40]. In one instance, Axie Infinity, a platform that promised users real-life returns for playing, i.e., a "play-to-earn" game, suffered a devastating hack that lost over

WEBSCI '24, May 21–24, 2024, Stuttgart, Germany

ACM ISBN 979-8-4007-0334-8/24/05
https://doi.org/10.1145/3614419.3644026

$620 million: 40% of the platform's users were from the Philippines, many of whom had invested their life savings into playing the game in hopes of earning high returns from the platform [7]. The hack left many users in crippling debt with little to no reparations [21]. This Ronin Bridge hack of Axie Infinity, one of the largest security breaches in the crypto space, not only had a profound impact on the game's ecosystem but also sparked extensive discussions across social media platforms, exemplifying how significant real-world events in the cryptocurrency world can rapidly translate into widespread conversations and sentiment shifts in the digital realm.

As the communities behind various blockchain projects grow, there is increasing activity on social media related to these topics. These discussions can offer valuable insights on various topics in the blockchain world and possibly even forecast events that take place in real life. However, the large volume of online activity also incurs difficulties in finding valuable information due to the varying quality of contributions and the sheer volume of the data. We addressed this issue by leveraging semantic-aware embeddings and a network-based approach to parse the data.

Social media platforms such as Twitter (now called X) allow for many types of analysis thanks to the textual nature of the posts. The tweets can be used to detect the topic of different groups of users, as well as approximate the sentiment across different topics using pre-trained language models such as BERT [8] or RoBERTa [26]. In some cases, these insights can provide signals pertaining to certain events' occurrence [12]. Similarly, the sentiment analysis of tweets related to blockchain technologies and decentralized finance (DeFi) can be leveraged to understand the broad user base sentiment towards various projects or aspects of blockchain technologies.

In this work, we conduct a large-scale analysis of Twitter posts related to cryptocurrencies collected to understand the possible correlations between signals in social media and real-life events and the conversation patterns found in Crypto Twitter[1] We used Twitter's v1 API [38] to collect tweets that contain cryptocurrency-related keywords and built a large dataset of around 40 million tweets. Our data was collected immediately following the height of cryptocurrency bridge hack incidents between November 9, 2022, and November 23, 2022. It is also worth noting that during this period, the FTX bankruptcy was made public [36], which was a formative period in public sentiment toward cryptocurrency. Thus, we find many mentions related to this incident in our dataset. However, instead of focusing on a specific attack vector or an incident, we aim to do a broader analysis of real-life events in cryptocurrency and possible signals that can be detected from social media. We achieve this by conducting an overall analysis of Crypto Twitter, scanning the broad sphere to find correlations between various real-life events and the users' reactions.

Our contributions are as follows:

- Topic modeling and sentiment analysis of tweets specific to blockchain projects or issues.
- A study of the correlation between sentiment scores of different tweet groups and real-life blockchain hacks, scams, and incidents.
- A graph-based analysis of tweet-level and user-level interactions in Crypto Twitter to further understand the social dynamics of the space.

[1]Crypto Twitter refers to the vibrant and active online community on the Twitter platform, where enthusiasts, investors, developers, and thought leaders in the cryptocurrency space congregate to share news, opinions, analyses, and insights about various aspects of cryptocurrencies and blockchain technology.

2 RELATED WORK

In navigating the intricate landscape of cryptocurrency and its social media dynamics, it is essential to explore the existing body of research, which sheds light on the multifaceted interactions between digital currency markets and online communities and the broader implications of these relationships on market sentiment and user behavior. Building upon this foundation, Guggenberger et al. [14] have significantly contributed to the literature on blockchain security, noting a marked increase in research in this area since 2013. Complementing this, Vokerla et al. [39] discusses a more focused analysis, characterizing ten prevalent methods of exploits across five distinct applications of blockchain technology.

Mirtaheri et al. [27] used a random forest classifier to predict occurrences of cryptocurrency pump-and-dump scams using social media signals. Since pump-and-dump scams encourage the purchase of assets to inflate prices artificially, the authors posit that sudden increases in chatter surrounding a particular project may indicate a pump-and-dump scheme in progress. The resulting classifier predicts these schemes across social media platforms, such as Twitter and Telegram, with an average accuracy of 74% ± 8%. The authors also identify an increase in bot accounts for generating artificial posts while a scheme is underway. However, unlike Mirtaheri et al. [27], which only focused on pump-and-dump schemes, we analyze various security flaws, hacks, and sentiments related to major events in the crypto space.

Saad et al. [33] analyzed the frequency of Google searches using terms related to a type of attack known as *cryptojacking*. The findings reveal a positive correlation between search term frequency and a large-scale attack of the indicated character. The authors discuss attack vectors in theory and in relation to high-profile attacks such as the one that bankrupted the cryptocurrency exchange Mt. Gox in 2013 [6].

Previous work has also used social media data to find the correlation between user sentiments and the price of different cryptocurrencies. Huang et al. [17] applied a long-short term memory (LSTM) based model to analyze Chinese social media sentiment and accurately predict cryptocurrency price fluctuations. The model used a lexicographical approach to creating a custom dictionary of cryptocurrency terms and achieved performance increases in precision and recall over a standard auto-regression model for time-series prediction. Pano and Kashef [29] used tweets related to Bitcoin to calculate the VADER sentiment scores [20] and obtained the score's correlation with Bitcoin price data. Lamon et al. [23] leveraged Twitter sentiment analysis to build a price prediction model for Bitcoin, Ethereum, and Litecoin. The authors find that the model can successfully predict large margins of price change, although more specific predictions are more difficult. Kraaijeveld and De Smedt [22] employed a similar analysis using Twitter data against cryptocurrency prices, concluding that Twitter data are a good predictor for some cryptocurrencies, such as Bitcoin, Ethereum, and Litecoin. The authors also note the heavy presence of bot accounts in the

blockchain Twitter space, noting that 1∼14% of the collected data are likely spam.

Finally, the work most similar to ours is presented by Linton et al. [24] and Park and Lee [30].

Linton et al. [24] used a modified form of Latent Dirichlet Allocation (LDA) capable of preserving time-series relationships in their study of topic modeling of discussion forum posts retrieved from bitcointalk.org, a platform for discussing cryptocurrencies and crypto-related events. Their findings, of particular interest for our experiments, support the identification of cryptocurrency-related events like suspected attacks on exchanges through clustering. It should be noted that their work finds that chatter related to all attacks was joined in a single cluster rather than one per attack, suggesting the need for further research into models that enhance attack-level granularity. In our work, we consider Twitter as the data source and use topic modeling and sentiment analysis to draw correlations to real-life events.

Park and Lee [30] analyze different types of networks in Crypto Twitter to identify the correlation in project ratings of different cryptocurrencies. Specifically, the authors use different relationships between tweets, such as follow-follow, reply-mention, and quote, of the official accounts of cryptocurrency projects to construct networks and draw a correlation to the Weiss Rating, which evaluates the financial strength and investment risk of various cryptocurrency projects aimed at guiding investors with independent and unbiased evaluation of the potential and stability of these projects. The authors find that the follow-follow network correlates highly with the Weiss Rating. However, this work focuses on the network structure of only *official accounts*, while our work considers the Twitter networks of *all* users we have in the dataset.

3 METHODOLOGY

In this section, we describe the methods used to collect our dataset and the analysis methods.

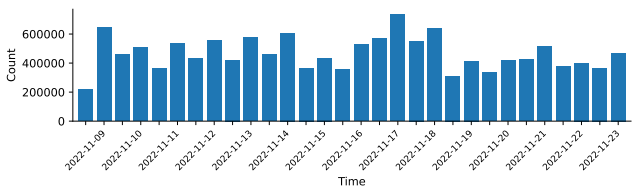

Figure 1: Volume of collected tweets from November 09, 2022, to November 23, 2022, binned in 12-hour intervals.

3.1 Data Collection

We manually curated a list of keywords related to cryptocurrency and security, ranging as broadly as "crypto" or "blockchain" to specific technologies such as "zk-rollup" or "layer-2." We also included security-related terms such as "breach" or "hack" to get tweets related to cryptocurrency hacks. In total, we collected our dataset using a list of 199 keywords. The full list of keywords is available in our anonymized repository. The data collection ran over 14 days, from November 09, 2022, to November 23, 2022. The final dataset consists of 39,828,572 unique tweets.

Field	Description
id	Unique ID of the tweet
timestamp_ms	Timestamp of the tweet
user.id	User ID
full_text	Main text of the tweet
quote_count	Number of quotes on the tweet
reply_count	Number of replies on the tweet
retweet_count	Number of retweets on the tweet
favorite_count	Number of "favorites" on the tweet
quoted_status.id	ID of tweet being quoted
quoted_status.user.id	User ID of the tweet being quoted
in_reply_to_status_id	ID of the tweet being replied to
in_reply_to_user_id	User ID of the tweet being replied to

Table 1: Features Used in Analysis

We used Twitter API v1 [38] to scan and scrape tweets that contain the selected keywords in real-time, maintaining our databases using Elasticsearch [10]. If a new tweet contains one or more keywords, it is added to the Elasticsearch index. The volume of tweets during our collection period can be seen in Figure 1. The complete index contains 130 features about each tweet. We considered a subset of these features, such as the text or reply/quote structure, in our analysis. Some of the features used in our work are shown in Table 1.

3.2 Data Sanitization

We found that some of the tweets in the dataset do not discuss any cryptocurrency-related topics, such as discussing just about "security" unrelated to the domain we are exploring. However, in the later stages of our analysis, the cluster-based and graph-based approaches can identify such tweets and group them accordingly.

In the context of Crypto Twitter, a *spam* tweet refers to an unsolicited and often irrelevant message, typically intended to promote certain cryptocurrencies, ICOs, or other crypto-related schemes. These tweets are usually characterized by their repetitive nature, excessive hashtags, and links to external sites. They might also include misleading information, false promises of high returns, or attempts to impersonate well-known figures in the crypto community. The primary goal of spam tweets in the crypto context is often to manipulate market sentiment, spread misinformation, or engage in fraudulent activities such as phishing or scams. Unsurprisingly, we found many such *spam* tweets present in our initial dataset of 40 million tweets.

However, we also note that some of these tweets could have been made by genuine users who share a lot of enthusiasm about the topic or are after an "Airdrop" of a certain crypto token [2]. This introduces some difficulties regarding our approach to definitively identifying spam tweets. With this in mind, we only removed tweets verified to be linked with a scam attempt. For example, we found that millions of tweets in our dataset contain the same set of sentences "Uniswap is being exploited by this dude. Why is nobody talking about this?...". These tweets appear to have been made by a coordinated bot attempt to push a phishing campaign [37], and were thus removed from our dataset.

After manually filtering the tweets by matching commonly found spam patterns that we were able to verify, we ended with 20,911,310 tweets in the final dataset for the analysis.

In addition to filtering for spam tweets, we replace user mentions with *user* and any external links to *http*, which allows us to better capture the style of tweets found in the dataset instead of specific mentions of individual users or links.

3.3 Sentence Embedding

We used a publicly available pre-trained Sentence BERT (SBERT) [32] model from HuggingFace [18] to encode the collected tweets into a latent space. SBERT is a BERT-based model that has been shown to outperform BERT in both *SentEval* and *Semantic Text Similarity (STS)* tasks [32]. External URLs and user mentions were replaced with the keywords "http," and "user," respectively, to avoid overfitting on terms that appear in usernames and URLs.

3.4 Sentiment Analysis

We used a publicly available RoBERTa [26] model fine-tuned on the TweetEval sentiment analysis task benchmark [3] to generate the sentiment scores used in our analysis. The publicly available weights from HuggingFace [19] were used without any additional fine-tuning. Unlike Sentence-BERT which is fine-tuned using a contrastive objective to predict the semantic similarity of sentences, this RoBERTa model was fine-tuned with a classification layer for predicting sentiment labels corresponding to the input text. The sentences are fed to the model to yield three types of sentiments [*positive*, *neutral*, and *negative*] and the confidence score associated with each label.

3.5 Tweet Clustering

In order to cluster the tweets by their semantic similarity, we used the SBERT embeddings [32]. We reduced the dimensionality of the embeddings by training and applying a linear autoencoder, then performed K-means clustering to group the tweets by their similarity in the latent space. We then used a term-frequency inverse-document-frequency (TF-IDF) matrix to extract the topics for each group. Even after dividing the tweets into different clusters, we faced a very large number of tweets to examine in each cluster. Therefore, to aid in this process, we implemented a dashboard (see Figure 2) to search each cluster of tweets effectively. Given a certain query string, this dashboard embeds the query into a SBERT embedding and searches for tweets with similar embeddings (nearest-neighbors by cosine similarity). Using the keywords observed in our TF-IDF analysis, we were able to take a deeper dive into the collected tweets by manually building queries using keywords contained in our clusters.

In order to get a cluster-specific set of topics, we excluded the top 10 terms found in the overall dataset from the top terms of each cluster.

We extracted the sentiment score of each tweet's raw text using the RoBERTa model fine-tuned on the TweetEval sentiment benchmark. To understand the sentiment change over time, we analyzed the keywords associated with each type of sentiment. We then grouped the tweets in each cluster by their tagged sentiment type and conducted TF-IDF analysis on each sentiment group separately.

3.6 Network-Based Analysis

In addition to clustering tweets using a semantics lens, we examined the dataset at the thread and user levels by constructing graphs of tweets. For this purpose, the `reply_to` and `quoted_status` fields of the tweets were utilized that contain the user ID and the tweet ID if it exists. We chose these fields because we wanted to model the user-to-user and tweet-to-tweet interactions to understand the conversation dynamics better. To achieve this, we constructed different graphical views of user interaction and tweets using these fields and analyzed their reply and quote interactions. We thus constructed and analyzed four types of graphs: tweet-quote, tweet-reply, user-quote, and user-reply.

We considered two types of components to analyze, namely *weakly connected components* (WCC) and *strongly connected components* (SCC). A weakly connected component is a subgraph where each node is reachable to every other node *disregarding* the direction of the edges. A strongly connected component is a subgraph where each node is reachable to every other node *while considering* the direction of the edges. Our analysis involving graphs was implemented using Python networkx library [16] and Gephi visualization [4].

4 ANALYSIS RESULTS

The motivation behind our analysis is twofold. We wish to understand the types of topics discussed in Crypto Twitter, as well as the conversation dynamic present in it. Thus, we performed two different types of analysis on our dataset. In the first analysis, we grouped the tweets and analyzed the topics and the sentiments associated with each group. This method focuses on understanding the types of tweets that exist in Crypto Twitter, such as the topics and the associated sentiment values. The second analysis is graph-based, in which we constructed a tweet-level graph based on the *reply* and *quote* relationships of the tweets as well as a user-level graph with the same *reply* and *quote* relationships. This method aims to better understand the dynamics of conversation within Crypto Twitter.

4.1 Topic Clustering

ID	Characterization	Top 10 Keywords
	Overall	user, http, pump, just, signal, happen, crypto, wallstreetbets, event, kucoin
1	Crypto Conspirators	user, pump, http, signal, just, event, happen, wallstreetbets, kucoin, big
2	Meta Crypto Twitter	user, crypto, http, promote, roll, token, price, 000, binance, security
3	Crypto Observers	user, http, crypto, v2, rollup, address, tokens, claiming, compatible, evm
4	Crypto Commenters	roll, crypto, sushi, user, project, security, good, try, http, bridge
5	Crypto Doubters	user, http, pump, crypto, just, signal, kucoin, happen, event, wallstreetbets
6	Interested Investors	promote, user, crypto, price, roll, http, btc, binance, eth, bitcoin

Table 2: Top 10 Words per Cluster

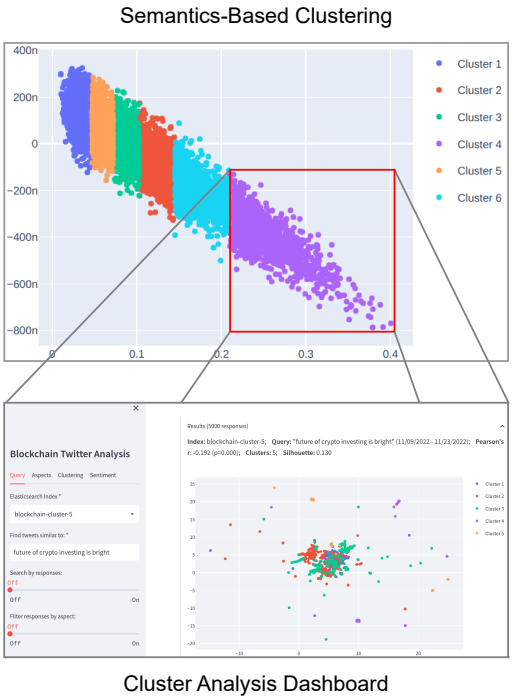

Figure 2: Pipeline of the two-stage topic clustering process. 1) The BERT embeddings are used to form clusters of the entire dataset using K-means. 2) Each cluster is then examined manually via a web dashboard for further analysis.

The keywords associated with positive and negative sentiment in each cluster can be seen in Table 3 and Table 4, respectively. Similar to how the clusters' topics were found, we excluded the top 10 terms of each sentiment at the dataset level to reduce the number of overlapping terms across clusters.

In order to capture the keywords of each cluster, we fitted a TF-IDF matrix over the entire document (every tweet) and per cluster, then selected the top 10 keywords for the overall document and each cluster. This process removes the most frequent keywords at the document level from each cluster, allowing for more granular insight into each cluster. The document-level keywords are marked as *overall* at the top row of each table. However, some overlap between the keywords of different clusters can still be observed. These can be interpreted as the overall trend captured in our dataset that our keyword filtering process could not remove.

The overall sentiment change over time in each cluster can be seen in Figure 3.

Cluster 1: Primarily comprises topics related to *group activity* or *price manipulation*. It contains tweets advertising private price manipulation groups, similar to the phenomenon noted by [28]. The two main targets of these price manipulation tactics are Binance and Kucoin, two popular cryptocurrency exchanges. An in-depth analysis of characteristic tweets reveals multiple "pump signals" announcements on Binance and Kucoin by groups attempting to perpetrate pump-and-dump schemes. The amount of identical or near-identical tweets makes this a noteworthy example of spam within the dataset relevant to mass crypto-exploitation and potentially harmful trading activity. This observation is supported by several of the top 10 words per cluster seen in Table 2. Top terms associated with positive sentiment are related to announcements of events, such as "massive," "just," or "announced". We also found that "binance" appears as a positive term, while "ftx" appears as a negative term. Most of the negative sentiment associated with FTX can be linked to the event where the head of Binance decided to stop their support for FTX [36]. We found tweets expressing distrust towards FTX even before the public announcement of the event, and threads discussing the news of Binance walking away from a deal with FTX, expressing their distrust towards FTX and its founder. We also found negative sentiment towards not just FTX but also people who supported FTX, with many tweets blaming the victims for trusting FTX.

ID	Top 10 Keywords
Overall	crypto, project, binance, token, signal, blockchain, pump, happen, good, big
1	wallstreetbets, massive, announced, event, really, airdrop, kucoin, join, best, awesome
2	roll, love, security, best, great, kinderinu, world, bridge, join, 10
3	roll, best, security, love, amazing, join, great, nft, like, team
4	roll, sushi, nice, want, love, great, dodo, let, bridge, security
5	best, join, wallstreetbets, love, massive, nft, airdrop, awesome, announced, security
6	roll, sushi, security, kinderinu, love, new, bridge, nfts, best, things

Table 3: Top 10 Terms Associated With Positive Sentiment

ID	Top 10 Keywords
Overall	crypto, security, people, roll, like, hack, ftx, binance, bridge, money
1	know, uniswap, want, time, blockchain, social, think, going, need, market
2	know, going, shit, token, threat, want, bad, time, scam, think
3	know, want, going, time, think, threat, token, need, bad, national
4	sushi, bad, shit, malicious, scam, fuck, losers, dumps, hour, trade
5	know, want, going, time, blockchain, think, national, need, social, token
6	sushi, going, bad, shit, 2022, scam, malicious, know, vulnerability, fuck

Table 4: Top 10 Terms Associated With Negative Sentiment

Cluster 2: Unlike in cluster 1, cluster 2's top 10 keywords were less useful in determining the overall topic of the cluster, because several words, like 'security,' 'roll,' and 'bridge,' are associated with many tweets unrelated to cryptocurrency and DeFi. For example, using the keyword "crypto" did not initially reveal a unifying topic of discussion, as tweets ranged from short statements like *"crypto*

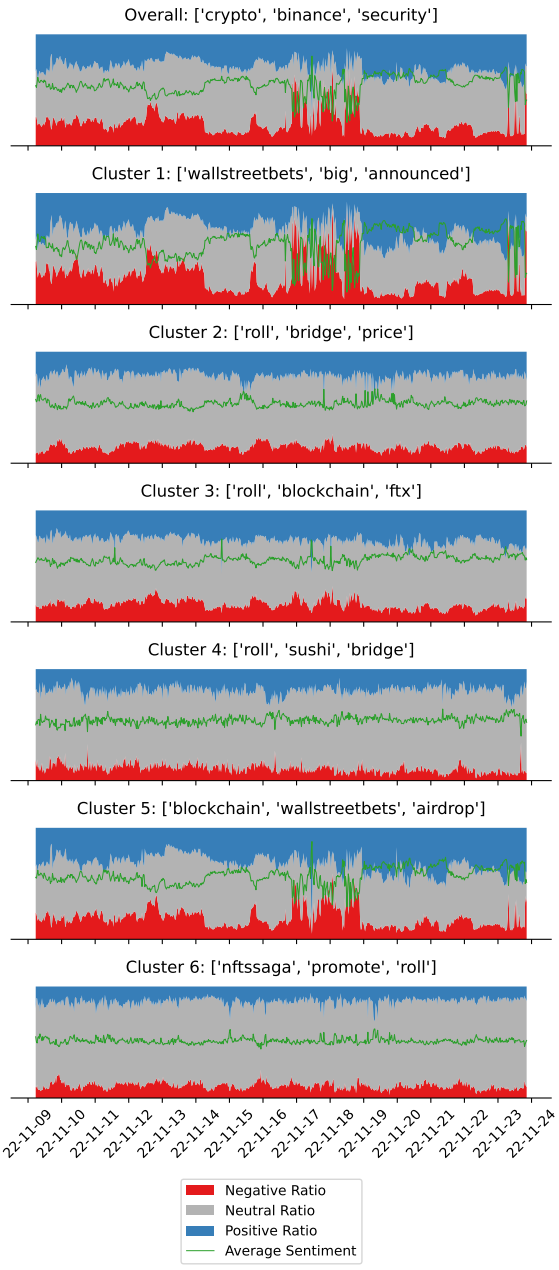

Figure 3: Change in sentiment score over time for each cluster with top-3 cluster-specific terms.
blue (—): ratio of positive sentiment, red (—): ratio of negative sentiment, gray (—): ratio of neutral sentiment, green (—): average sentiment

is the future" to *"Why should someone use crypto"*. However, using our two-part clustering method, we found that one sub-cluster of cluster 2 contained many tweets directed toward Elon Musk, who has often weighed in on crypto-related events and promoted various coins. Since Elon Musk is an influential user within the DeFi space, there were many tweets asking for his opinion on specific projects or requesting promotion for tokens.

We found keywords such as "project," "blockchain," and "token" associated with positive sentiment, while keywords such as "hack," "roll," "bridge" are associated with negative sentiment. Many of the positive sentiments were originated from tweets expressing their gratitude towards developers. We found many tweets similar to "Great project" or "Thank you for this project" in positive sentiments. Looking into the negative sentiments, we found tweets expressing their distrust in "bridge hacks" and the security features of many bridge protocols. We also found many specific mentions to the "Ronin Bridge", a project related to Axie Infinity, which was hacked in July [21] with a loss of over 600 million dollars.

Cluster 3: Here, we found chatter regarding cryptocurrency exchanges, with users expressing distrust in these institutions and, less commonly, optimism regarding the trajectory of DeFi. Searches using the keywords "binance" and "FTX" revealed a general mistrust in exchanges. Interestingly, mistrust of centralized exchanges (CEXes), such as FTX, does not always translate to mistrust in cryptocurrency as a whole, with many users advocating for a shift away from CEXs toward more decentralized methods of digital currency storage, such as offline wallets. These opinions are unsurprising in light of the FTX incident [36], which occurred during data collection and whose effects are still felt within the market months later.

We found many overlapping terms for positive and negative sentiment. Terms such as "security" and "roll " appear in both sentiments, suggesting that multiple sub-clusters in this cluster express their trust and distrust of centralized exchanges regarding these topics. We also find that "binance" appears as one of the top positive terms, while "ftx" appears among the negative terms. Looking deeper, we find that many of the tweets with a positive sentiment that mentions Binance are thanking the CEO of Binance. In contrast, FTX is mostly mentioned in a negative context, especially after the official bankruptcy announcement.

Cluster 4: This cluster contained the most syntactically distinct tweets, with most only two or three words long. Common tweets within this cluster were *"crypto fan"* and *"to the moon"*, a phrase typically used to signify one's hope that a project will take off and rapidly accrue investors. These quick comments are in response to other tweets intended to generate engagement, and thus represent a subset of ordinary users within Crypto Twitter space rather than influencers or official project social media accounts.

As pointed out in our earlier analysis, cluster 4 comprises mostly short-form tweets. Many positive and negative keywords are either adjectives or verbs with strong sentiment attached, such as "nice," "love," "want" for positive sentiment, "malicious," "hack," and "bad" for negative sentiment. We also note that the term "security" appears in both positive and negative sentiment, suggesting that tweets in this cluster also discuss the security aspect of various projects in both positive and negative contexts.

Cluster 5: This cluster contained tweets related to the security of various blockchain projects, with users tweeting directly at the official accounts of various projects to ask questions regarding security and trust. This characterization is upheld when querying for "binance." Many tweets discuss transparency and the duties of

exchanges to their customers. Tweets supporting increased transparency and regulations on blockchain trading activities show the community's growing interest in ensuring safety and reliability in projects that handle large amounts of money.

In our earlier analysis of the topics, we found that this cluster is the least coherent in terms of the topics and contains many advertisements for different blockchain assets/protocols. As a result of this, we did not find any specific mentions of projects or exchanges except for Binance, which was a very engaging topic at the time of our data collection. Our sentiment analysis supports the finding that most positive sentiments are associated with advertisements, with keywords such as "just," "join," and "best" appearing in the top positive terms.

Cluster 6: In this cluster, we found a multitude of tweets discussing various tokens. Notably, this differs from the promotional activity seen in cluster 1. Unlike the spam advertisements that dominate the advertisement space of that cluster, these tweets are unique affirmations or criticisms by users often tweeted directly at official project accounts. Querying this cluster using the keyword 'token' yielded tweets such as *"I trust in WorldCupinu token"* and *"$CREAM token still flying."* This shows that despite the various scam activities in the crypto space, some users have genuine faith in ongoing projects.

We found that cluster 6 has the highest number of specific names mentioned in the blockchain space associated with positive sentiments, such as "binance," "kinderinu," and "sushi". Sushi is likely in reference to the Sushi Swap decentralized exchange, while Kinderinu (*kinderinu.io*) is a type of "meme" token similar to Dogecoin (*dogecoin.com*), which gained popularity among some investors. Looking deeper, we found that many tweets that mention the "kinderinu" project have the same content, similar to the Uniswap liquidity spam tweets mentioned in the Methodology section. Some of the tweets in this cluster were made under a coordinated attempt, although we could not find a source to verify.

4.2 Tweet Graph

The tweet graph is a directed graph in which each node is a tweet. A directed edge from tweet A to tweet B indicates that tweet A is a reply or quote to tweet B. We filtered for tweets that contained valid text to gather a subset of the dataset, which excluded retweets and tweets that refer to another tweet with a quote/reply. Consequently, each node in our graph is associated with its respective text content. Additionally, we computed RoBERTa sentiment scores for these texts. Given that each tweet (node) can quote or reply to at most one other tweet, resulting in at most one outgoing edge per node, our tweet graph tends to be sparse. Therefore, our analysis primarily concentrated on WCCs, as these represent more significant clusters or groups within the reply and quote dynamics of the graph.

The statistics of the generated graph can be seen in Table 5. As seen in the degree distribution, the quoted graph is much sparser than the reply graph. This indicates that users in our dataset communicate more frequently using replies instead of quotes. We identified the ten largest WCCs for each version of the node relationship considered and analyzed the types of tweets present in each of the largest WCCs. Examples of the tweet-reply and tweet-quote graphs

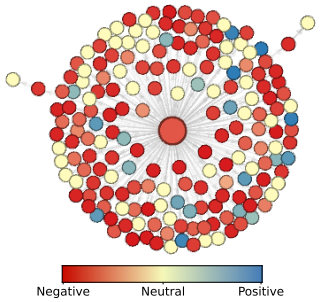

Figure 4: Largest WCC of the Tweet-Quote graph.
$|V| = 188, |E| = 187$.

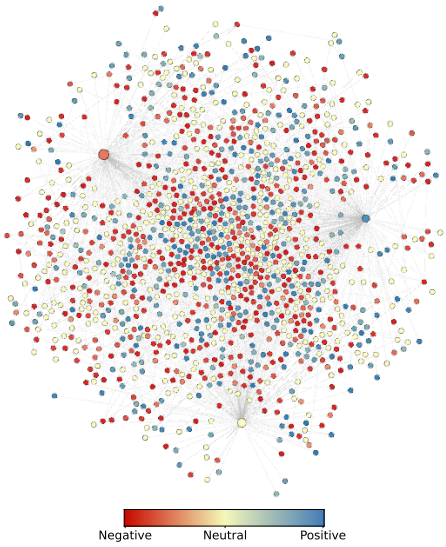

Figure 5: Seventh largest WCC of the Tweet-Reply graph.
$|V| = 1282, |E| = 1281$.

can be seen in Figure 4 and Figure 5. The nodes are colored by their sentiment and sized by their interaction count.

- Nodes with shades of red (●) have negative RoBERTa sentiment, darker color being stronger.
- Nodes with yellow (●) have neutral RoBERTa sentiment.
- Nodes with shades of blue (●) have positive RoBERTa sentiment, darker color being stronger.
- Nodes are sized by the *interaction count*, which is a sum of `retweet_count`, `favorite_count`, `reply_count` and `quote_count`.

Tweet-Quote Graph: The tweets in the largest WCC (WCC-1) discuss an incident from Crypto.com, a centralized cryptocurrency exchange, in which the exchange had accidentally misplaced its funds [34]. The interactions appear from genuine users, many expressing their confusion and concerns about the incident. A snapshot of this interaction can be seen in Figure 4. The central tweet in the sixth largest WCC (WCC-6) discusses that some cryptocurrency exchanges appear to be swapping funds to build a fake *snapshot* of their reserves. Further inspection reveals that this tweet was made

Graph Type	# Nodes	# Edges	Degree			In-Degree			Out-Degree		
			Mean/std	Median	Max	Mean/std	Median	Max	Mean/std	Median	Max
Tweet-Quote	44,300	26,255	1.185/1.893	1	184	0.592/0.491	1	1	0.592/1.998	0	184
Tweet-Reply	1,591,247	1,034,017	1.299/5.952	1	6,391	0.650/0.477	1	1	0.650/5.978	0	6,391
User-Quote	1,640,251	2,520,608	3.073/144.905	1	59,825	1.537/6.068	1	3,736	1.537/144.788	0	59,825
User-Reply	2,340,960	7,298,105	6.235/75.741	1	49,486	3.118/52.859	1	13,051	3.118/52.453	1	49,481

Table 5: Statistics of the Tweet and User Reply/Quote Graphs

by a Crypto Twitter influencer, referencing a scandal where several centralized exchanges were suspected to lack the liquidity they claim [35]. All the other WCCs (WCC-2, WCC-3, WCC-4, WCC-5, WCC-7, WCC-8, WCC-9, and WCC-10) contain mostly bot attempts, and a few notable bot activities can be characterized as follows:

- Tweets attempting to shine a positive light on the Terra-Luna project after its collapse in May [25].
- Fake airdrop scams that advertise a token from a project that had not been released or advertised, leading to a third-party website for phishing [1]. For instance, our dataset had many mentions of zksync (*zksync.io*), a platform known for its airdrops, and the users linking a URL to a different (phishing) website.

Tweet-Reply Graph: The largest WCC (WCC-1) in this graph mostly comprises tweets with the same two hashtags about a token called "Leonicorn." At first glance, this appears to be a group of spam tweets. However, many of the user accounts behind these tweets appear legitimate, with up to hundreds of followers and previous activity on Twitter. We conclude that this may be an example of a coordinated attempt by users to bring certain hashtags to the *trending* list of hashtags on Twitter. While the accounts appear to be regular Twitter users, this activity reveals another example of manipulation on Crypto Twitter.

The third, fourth, and tenth largest WCCs (WCC-3, WCC-4, WCC-10) reveal another interesting insight into spam activity on Crypto Twitter. There is a reply chain composed almost entirely of spam bots with regular users scattered throughout. The bot accounts appear to reply to seemingly randomly chosen users who discuss cryptocurrency-related topics. We also found instances where bots advertising different scams reply to each other in a chain that starts with a tweet from a genuine user. Many of these bots advertise an "exploit" the user can benefit from, such as Maximal Extractable Value (MEV) scripts, which exploit the ability to influence the order of transactions in a block to gain advantages, or an upcoming "strategy" such as a rug pull on an asset, where the developers of a cryptocurrency project suddenly withdraw all their funds from the liquidity pool or project wallet, abandoning the project and leaving investors with worthless tokens or assets. We found several instances where regular users attempt to respond to the bot tweets, only to be swarmed by further tweets advertising a different "exploit."

The sixth largest WCC (WCC-6) is centered around a tweet announcing a giveaway event from Binance's official account. While this tweet is legitimate, most replies are tweets from low-follower users that link to an external website in a likely scam attempt, which reveals another pattern of spam bots on Crypto Twitter. Unlike other spam tweets observed in the dataset, these tweets only mention the official Binance account. Instead of mentioning multiple users with a description of the "exploit," these tweets use a few words such as *"Interesting"* or *"Important."* We conclude that these tweets attempt to lead possible victims into visiting the link out of curiosity and then fall for the scam. Furthermore, we found that many of these links lead to deleted videos on YouTube , which was a popular format for sharing the "tutorials" on these "exploit" scams.

The second and fifth largest WCCs (WCC-2, WCC-5) appear to stem from Binance's founder and former CEO Changpeng Zhao. WCC-2 is centered around a tweet from Zhao, which says *"I think I am a poor speaker. I stutter all the time."* This tweet is met by a majority of positive support from the community, with the users cheering him on and expressing their faith in him. WCC-5 is centered around Zhao's tweet *"Rebuilding starts, to the moon."* announcing a rebuilding of the crypto community after the FTX incident. This tweet is met with mixed sentiment, with many users expressing distrust in the centralized exchange model.

Interestingly, when Binance's official account tweets, those are met with much more negative responses. The seventh largest WCC (WCC-7) contains three tweets from Binance's official account, which appear to be from a thread the account posted on its position in light of the FTX instance. Even though the tweets attempt to explain Binance's decision not to support FTX, many tweets express the users' discontent with Binance's actions and distrust of the platform. Figure 5 shows a visualization of WCC-7, where the tweets from Binance's official account are the red (●), blue (●), and yellow (●) larger nodes in the top left, center right, and center bottom of the diagram, respectively.

A tweet from Etherum's founder, Vitalik Buterin, is found to be the central node of the ninth largest WCC (WCC-9). Interstingly, while Vitalik's tweet mentions his doubts about the future of ZK-rollup, an overwhelming majority of the replies to this tweet are positive. We found that many of these replies are not related to the original tweet by Vitalik but are either promoting a likely scam or proposing ideas to him, many of which are classified as having a positive sentiment.

4.3 User Graph

We constructed two distinct user graphs: one based on *quotes* and the other on *replies*. These are also directed graphs, with nodes representing individual users and edges representing interaction between them. In the user-quote graph, a directed edge from user A to user B signifies that user B has quoted a tweet from user A. In the user-reply graph, a directed edge from user A to user B indicates that user B replied to a tweet from user A. Consequently, we filtered and created edges from only those instances that provide both. Unlike

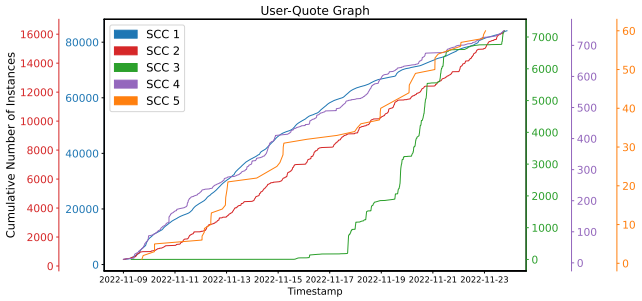

Figure 6: Change in Number of Data Points Involved with the SCC Nodes Over Time (User-Quote Graph)

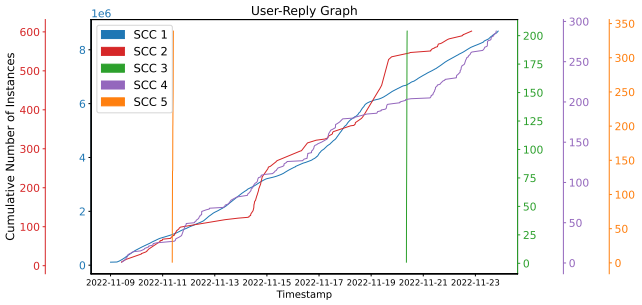

Figure 7: Change in Number of Data Points Involved with the SCC Nodes Over Time (User-Reply Graph)

the tweet graph, some nodes in these graphs feature both inbound and outbound edges, along with self-loops. Thus, we turned our attention to the SCCs to analyze significant groups within both the reply and quote graphs. This analysis reveals diverse discussions, spam bot networks, and engagement levels in the conversations.

Table 5 presents statistics for the generated graphs. Like the tweet graph, the user-reply graph exhibits higher density than the user-quote graph, suggesting a preference among users for replies as their interaction medium over quotes. We examined the five largest SCCs for both of the graphs.

User-Quote Graph: In the context of SCC-1, the largest component, discussions center around the event regarding the collapse of FTX [31]. This subnetwork is dominated by legitimate users, as evidenced by the high mean and median number of friends and followers (see Table 6).

SCC-2, the second-largest component, is smaller, with 25 nodes and 105 nodes reachable from it. Intriguingly, nearly all nodes, including reachable ones, are identified as spam bot users responsible for disseminating over 16,000 spam tweets, primarily glorifying AdsCoin (*https://adsexchange.io*). Initial manual examination of tweets confirmed their spam bot nature, and subsequent analyses discussed in later sections further support this claim.

The third-largest component (SCC-3), encompassing 21 nodes with 28 reachable nodes, unexpectedly diverges from crypto-related discussions. The dialogue here revolves around the "Save Kashmir movement" in India. However, since our dataset is filtered with keywords such as attacks, security, etc., relevant to crypto and non-crypto contexts, these data inadvertently entered the dataset, albeit unrelated to the crypto domain.

On a smaller scale, SCC-4, the fourth-largest component, comprising only 11 nodes, has a relatively larger number of reachable nodes (358). In this subnetwork, individuals associated with Cosmos (*cosmos.network*), a cryptocurrency powering a network of blockchains that can communicate with each other, engage in diverse topics. Unlike a singular discussion focus, Cosmos serves as the unifying element bringing nodes together in this subnetwork.

SCC-5, the smallest component with seven and four reachable nodes, lacks a specific central theme. Instead, it presents a microcosm of a social network where a small group of connected individuals interact on a range of random topics by quoting each other's tweets.

User-Reply Graph: The largest Strongly Connected Component (SCC-1) within the user-reply graph, comprising a substantial number of nodes (231,841), is centered around Binance (*www.binance.com*), the world's largest cryptocurrency exchange by trading volume. The majority of accounts within this component are associated with Binance, engaging in discussions on various topics, such as the decision not to pursue the potential acquisition of FTX and attending the Fintech summit in Indonesia. Although this subgraph is not entirely free from bots, the predominant user base consists of legitimate participants discussing diverse aspects primarily related to Binance.

Interestingly, the second and fourth largest SCCs (SCC-2 and SCC-4) focused on a power outage in Nigeria in November 2022. Tweets with the highest engagement in SCC-2 originate from the official account of Ikeja Electric (*www.ikejaelectric.com*), Nigeria's largest Electricity Distribution Company, providing updates and services during the outage. Similarly, in SCC-4, discussions revolve around another company, Abuja Electricity Distribution Company (*www.abujaelectricity.com*). Like the discussions involving SCC-2, these are also focused on updates and services during that power outage. Although these discussions are not directly related to crypto, Ikeja Electric does mention different tokens, such as energy tokens, which are relevant to their business. Notably, the mean followers count for nodes within SCC-2 is significantly lower than that of nodes reachable from SCC-2, deviating from the expected trend (see Table 7). Further investigation revealed that Elon Musk, who has a very large number of followers, is reachable from SCC-2. A more in-depth examination reveals the involvement of a spam bot connecting users discussing the power outage and those related to crypto. Elon Musk replied to a tweet from Chainlink, and a user connected with Chainlink interacted with the spam bot. This way, Elon Musk entered this network, and his substantial follower base caused the anomaly in the mean followers count. Other than this deviation, the subnetwork appears normal, although the context in focus is not very related to crypto. However, SCC-4 doesn't show such deviation and follows the normal trend.

The third largest SCC (SCC-3) is found to be centered around a Binance bot that posted numerous tweets about the CR7 NFT Giveaway in a short timeframe. This is the same bot network as one of the identified botnets in the tweet-reply graph.

The fifth-largest SCC (SCC-5) also constitutes a spam bot network, where users predominantly post hashtags and random spam

Observation	Mean of Friends		Median of Friends		Mean of Followers		Median of Followers	
	Within	Reachable	Within	Reachable	Within	Reachable	Within	Reachable
SCC-1	1375	1202	691	368	90352	7429	1019	198
SCC-2	422	404	207	290	365	271	234	107
SCC-3	3474	96	420	3	4080	81	84	27
SCC-4	2973	313	2226	4	17367	2558	11701	8
SCC-5	180	95	158	31	144	71	126	34

Table 6: Observations of Mean and Median (rounded to the nearest integer) Friends/Followers of the Nodes Inside SCCs and the Nodes that are Reachable from SCCs (User-Quote Graph)

Observation	Mean of Friends		Median of Friends		Mean of Followers		Median of Followers	
	Within	Reachable	Within	Reachable	Within	Reachable	Within	Reachable
SCC-1	448	296	39	20	26823	962	4	2
SCC-2	661	436	299	36	5875	24535	124	4
SCC-3	0	0	0	0	0	0	0	0
SCC-4	384	436	63	36	22487	24536	33804	4
SCC-5	180	95	158	31	144	71	126	34

Table 7: Observations of Mean and Median (rounded to the nearest integer) Friends/Followers of the Nodes Inside SCCs and the Nodes that are Reachable from SCCs (User-Reply Graph)

replies. Users here were active for a very short period and posted the tweets and replies in just one day. Manual verification of users' tweet and reply activities in this subnetwork confirmed their bot nature. It is worth noting that Twitter has suspended most of these users due to their spamming activities.

The temporal dynamics of user interactions within the SCCs shed light on the evolving nature of discussions over time. For each of the top 5 SCCs, we observed how the number of data points involved with the SCC nodes changes. To be more specific, for each of the top 5 SCCs, at each timestamp T, we counted the number of samples having timestamps less than or equal to T that are connected to/from any of the SCC nodes.

Figure 6 and Figure 7 illustrate tweet counts over distinct timestamps, thus offering a comprehensive view of the evolution of conversations. While most SCCs of the user-quote graph exhibit a non-linear pattern that mirrors the dynamic nature of real-life discussions with peaks and lulls, SCC-2 (the red (—) curve) deviates by showcasing a distinctive linear trend. This synchronous and unusual linearity aligns with the identified characteristics of SCC-2 as a network primarily composed of spam bot users. SCC-2 and SCC-5 of the user-reply graph also show distinctive patterns for spam bot networks. The temporal analysis captures the rhythm of discussions and provides a visual cue to discern patterns that distinguish bot-driven activity from organic, user-driven discourse.

Examining the number of friends and followers within the SCCs and nodes reachable from them accentuates the distinctions (see Tables 6 and 7). The friends and followers counts within most SCCs of the user-quote graph significantly exceed those of reachable nodes. Notably, SCC-2 follows a distinct trend where the counts within the SCC and the reachable nodes are small and comparable, with the reachable nodes even exhibiting a higher median friends count than the SCCs. This divergence provides additional evidence supporting the characterization of SCC-2 as a bot network. These particular statistics in the user-reply graph also follow a similar pattern. However, SCC-3 with 0 as all the counts indicate an anomaly supporting the previous claim that this subnetwork consists of bots. The anomalous small counts in SCC-5 and insignificant difference in the counts between the nodes within SCC and the nodes reachable from SCC indicate that this subnetwork is also a spam botnet, reinforcing the previously discussed insight about SCC-5. These friends and followers count within the SCCs and those reachable from them also reveal a concentration of interaction. In general, the counts within an SCC surpass those of reachable nodes, indicating that the primary discourse occurs among influential users. However, the reachable nodes are not negligible users; they serve as observers, sharing tweets through quote tweeting. More followers than friends signifies the natural trend for influential users.

5 CONCLUSION

We analyzed cryptocurrency-related tweets that we collected during an eventful period in the crypto world (the FTX event) using popular natural language processing techniques such as sentence embedding [32], topic modeling [13], sentiment analysis [3], and by building tweet-level and user-level graphs. In our cluster analysis, we successfully identified three types of tweets focused on specific aspects of cryptocurrency, such as investment, security, and price manipulation. Additionally, our analysis uncovered a significant presence of content creators and developers within these clusters, actively engaging in efforts to generate awareness and excitement for their projects. This engagement pattern offers a valuable resource for understanding user interactions and overlaps across various crypto projects. We also found that a correlation exists between real-life events in the crypto world and the sentiment captured through Twitter data, and that we can estimate the timeline of the FTX incident [36] using the sentiment spikes on some of our identified clusters. Interestingly, only some clusters in our dataset displayed a noticeable reaction to the incident, while other topics overshadowed the sentiment in other clusters. Although our dataset was gathered with a primary focus on cryptocurrency, our clustering results suggest that a more fine-tuned monitoring

of social media chatter, such as focusing on one specific project, could reveal more nuances in social media communities' sentiment towards them and offer valuable insights.

Our initial goal was to build a framework for real-time detection of incidents in crypto-twitter. However, evaluating a real-time detection framework is a difficult task due to the sheer volume of tweets in the space and the unpredictability of such incidents. Thus, we conducted analyses focusing on fraudulent activity on crypto-twitter, collected around the timeframe of the FTX incident. Given enough storage and computing resources, our work can be extended as a building block for such a real-time detection system.

While we initially identified that *at least* 8% of our data are spam tweets, there is a much higher bot activity than we could capture with a simple heuristic spam filter. Furthermore, our graph-based analysis shed light on the extent of bot activity on Crypto Twitter. A significant presence of bots was observed at the tweet and user levels, often clustering around popular threads. This prevalence of bots, coupled with the discovery of legitimate-looking accounts engaged in repetitive tweeting, underscores the challenge of identifying and managing spam content in Crypto Twitter. However, the clustering algorithm and graph components appear to place these bot tweets in a similar cluster, which allowed us to bypass most of the remaining spam tweets in our dataset.

Our research insights enhance the understanding of the social dynamics of the intricate landscape of Crypto Twitter, revealing the interplay between human and automated interactions. These insights also pave the way for future explorations into the digital ecosystems of cryptocurrency, highlighting the critical need for advanced, nuanced analytical tools to navigate and interpret the ever-evolving narratives within these vibrant online communities.

Resources: All the code used in our analysis is available online in our GitHub repository at https://github.com/blockchain-interoperability/blockchain-social-media. This includes the Tweet IDs of our dataset, our analysis workflow, and comprehensive documentation.

ACKNOWLEDGMENTS

The authors acknowledge the support from NSF IUCRC CRAFT Center research grant (CRAFT Grant #22008) for this research. The opinions expressed in this publication do not necessarily represent the views of NSF IUCRC CRAFT.

REFERENCES

[1] Joel Agbo. 2023. Airdrop Scams in Crypto and How to Avoid Them. https://www.coingecko.com/learn/airdrop-scams-crypto.
[2] Darcy WE Allen, Chris Berg, and Aaron M Lane. 2023. Why airdrop cryptocurrency tokens? *Journal of Business Research* 163 (2023), 113945.
[3] Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421* (2020).
[4] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154
[5] Huashan Chen, Marcus Pendleton, Laurent Njilla, and Shouhuai Xu. 2020. A survey on ethereum systems security: Vulnerabilities, attacks, and defenses. *ACM Computing Surveys (CSUR)* 53, 3 (2020), 1–43.
[6] Adrian Cheung, Eduardo Roca, and Jen-Je Su. 2015. Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. *Applied Economics* 47, 23 (2015), 2348–2358.
[7] Andrew R. Chow and Chad de Guzman / Manila. 2022. A Crypto Game Promised to Lift Filipinos Out of Poverty. Here's What Happened Instead. *Time* (July 2022). https://time.com/6199385/axie-infinity-crypto-game-philippines-debt
[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805
[9] Cheyenne DeVon. 2023. Crypto investors lost nearly $4 billion to hackers in 2022. *CNBC* (Feb. 2023). https://www.cnbc.com/2023/02/04/crypto-investors-lost-nearly-4-billion-dollars-to-hackers-in-2022.html
[10] BV Elasticsearch. 2018. Elasticsearch. *software], version* 6, 1 (2018).
[11] Gianni Fenu, Lodovica Marchesi, Michele Marchesi, and Roberto Tonelli. 2018. The ICO phenomenon and its relationships with ethereum smart contract environment. In *2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE)*. IEEE, 26–32.
[12] José Antonio García-Díaz, Óscar Apolinario-Arzube, José Medina-Moreira, Harry Luna-Aveiga, Katty Lagos-Ortiz, and Rafael Valencia-García. 2018. Sentiment Analysis on Tweets related to infectious diseases in South America. In *Proceedings of the Euro American Conference on Telematics and Information Systems*. 1–5.
[13] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).
[14] Tobias Guggenberger, Vincent Schlatt, Jonathan Schmid, and Nils Urbach. 2021. A Structured Overview of Attacks on Blockchain Systems. *PACIS* (2021).
[15] Aparna Gupta, Jyothsna Harithsa, and Oshani Seneviratne. 2020. A descriptive analysis of us initial coin offerings. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*. IEEE, 144–151.
[16] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, Gaël Varoquaux, Travis Vaught, and Jarrod Millman (Eds.). Pasadena, CA USA, 11 – 15.
[17] Xin Huang, Wenbin Zhang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. 2021. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. *Database Systems for Advanced Applications* 12683 (2021).
[18] HuggingFace. 2022. sentence-transformers/all-MiniLM-L6-v2. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2.
[19] HuggingFace. 2022. Twitter-roBERTa-base for Sentiment Analysis. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment.
[20] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.
[21] Steve Kaaru. 2022. Axie Infinity left thousands of Filipino players in debt, and now they are leaving the game. *CoinGeek* (July 2022). https://coingeek.com/axie-infinity-left-thousands-of-filipino-players-in-debt-and-now-they-are-leaving-the-game
[22] Olivier Kraaijeveld and Johannes De Smedt. 2020. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money* 65 (2020), 101188.
[23] Connor Lamon, Eric Nielsen, and Eric Redondo. 2017. Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev* 1, 3 (2017), 1–22.
[24] M. Linton, E. G. S. Teo, E. Bommes, C. Y. Chen, and Wolfgang Karl Härdle. 2017. *Dynamic Topic Modelling for Cryptocurrency Community Forums*. Springer, Berlin, Heidelberg, 355–372. https://doi.org/10.1007/978-3-662-54486-0_18
[25] Jiageng Liu, Igor Makarov, and Antoinette Schoar. 2023. *Anatomy of a Run: The Terra Luna Crash*. Technical Report. National Bureau of Economic Research.
[26] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL]
[27] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Fred Morstatter, Greg Ver Steeg, and Aram Galstyan. 2019. Identifying and Analyzing Cryptocurrency Manipulations in Social Media. *CoRR* abs/1902.03110 (2019). arXiv:1902.03110 http://arxiv.org/abs/1902.03110
[28] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, Maurizio Tesconi, and Emilio Ferrara. 2020. Charting the landscape of online cryptocurrency manipulation. *IEEE Access* 8 (2020), 113230–113245.
[29] Toni Pano and Rasha Kashef. 2020. A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19. *Big Data and Cognitive Computing* 4, 4 (2020), 33.
[30] Han Woo Park and Youngjoo Lee. 2019. How Are Twitter Activities Related to Top Cryptocurrencies' Performance? Evidence from Social Media Network and Sentiment Analysis. *Drustvena Istrazivanja* 28, 3 (2019), 435–460. https://doi.org/10.5559/di.28.3.04
[31] Nathan Reiff. 2023. The Collapse of FTX: What Went Wrong With the Crypto Exchange? https://www.investopedia.com/what-went-wrong-with-ftx-6828447.
[32] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 3982–3992. https://doi.org/10.18653/v1/D19-1410

[33] Muhammad Saad, Jeffrey Spaulding, Laurent Njilla, Charles Kamhoua, Sachin Shetty, DaeHun Nyang, and David Mohaisen. 2020. Exploring the Attack Surface of Blockchain: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials* 22, 3 (2020), 1977–2008. https://doi.org/10.1109/COMST.2020.2975999

[34] Arijit Sarkar. 2022. Crypto.Com Accidentally Sends 320k ETH to Gate.Io, Recovers Funds Days After. https://cointelegraph.com/news/crypto-com-accidentally-sends-320k-eth-to-gate-io-recovers-funds-days-after.

[35] Arijit Sarkar. 2022. Huobi and Gate.Io under Fire for Allegedly Sharing Snapshots Using Loaned Funds. https://cointelegraph.com/news/huobi-and-gate-io-under-fire-for-allegedly-sharing-snapshots-using-loaned-funds.

[36] MacKenzie Sigalos. 2022. Sam Bankman-Fried steps down as FTX CEO as his crypto exchange files for bankruptcy. *CNBC* (Nov. 2022). https://www.cnbc.com/2022/11/11/sam-bankman-frieds-cryptocurrency-exchange-ftx-files-for-bankruptcy.html

[37] Sebastian Sinclair. 2022. Uniswap Liquidity Pool Hit With Phishing Attack Totaling $3.5M in Ether. https://blockworks.co/news/uniswap-lp-hit-with-phishing-attack-totaling-3-5m-in-ether.

[38] TwitterAPI. 2021. Twitter API V1 Documentation. https://developer.twitter.com/en/docs/twitter-api/tools-and-libraries/v1

[39] Rahul Rao Vokerla, Bharanidharan Shanmugam, Sami Azam, Asif Karim, Friso De Boer, Mirjam Jonkman, and Fahad Faisal. 2019. An Overview of Blockchain Applications and Attacks. In *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. 1–6. https://doi.org/10.1109/ViTECoN.2019.8899450

[40] Skye Witley. 2023. Crypto Hack Lawsuits Rise as Theft Victims Try Untested Claims. *Bloomberg Law* (Jan. 2023). https://news.bloomberglaw.com/privacy-and-data-security/crypto-hack-lawsuits-rise-as-theft-victims-try-untested-claims