"Invest in crypto!": An analysis of investment scam advertisements found in Bitcointalk

Gilberto Atondo Siu
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
jga33@cam.ac.uk

Alice Hutchings
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
alice.hutchings@cl.cam.ac.uk

Marie Vasek
Department of Computer Science
University College London
London, United Kingdom
m.vasek@ucl.ac.uk

Tyler Moore
School of Cyber Studies & Tandy School of Computer Science
The University of Tulsa
Tulsa, Oklahoma, United States
tyler-moore@utulsa.edu

***Abstract*—This paper investigates the evolution of investment scam lures and scam-related keywords in the cryptocurrency online forum Bitcointalk over a period of 12 years. Our findings show a shift in scam-related keywords found within posts in the forum, where "Ponzi" was the most popular and most frequently mentioned in 2014 and 2018 and "HYIP" appeared more often in 2018 and 2021. We also identify that the financial principle is the tactic more likely to be used to lure people into investment scams from 2015 until 2017, coinciding with the period when "Ponzi" was the most commonly found keyword. This is followed by a transition to the authority and distraction principles from 2018 until 2022, which also coincides with the increase of popularity of "HYIP".**

We collect more than 17.8M posts from 399k threads from the forum from July 2010 until June 2022. Our longitudinal analysis shows the popularity transition between subforums and keywords across time. We design a categorisation criteria and annotate 4,218 posts from 2,630 threads based on it. We then use the annotated sample to train four machine learning statistical models. We use the best performing model to classify all 281k English-language threads into four categories: overt scams, potential scams, scam comments and not investment scam related. We analyze the frequency changes of scam-related threads across the 12 year period and observe that overt and potential scams peaked in 2015 and 2018 respectively. We see that potential scams also increased during the COVID-19 pandemic. We use heuristics to pinpoint the types of cryptocurrencies most frequently used within scam advertisements. Bitcoin is most commonly found in potential scams while Ethereum appears more often than other cryptocurrencies in overt scams. We use machine learning classifiers to identify the scam actor types behind the posts categorised as overt and potential scams. We also classify the type of lure used by scammers. Our results indicate that the time principle is not a tactic used as frequently as expected. Finally, we observe the influence of the pandemic in the strategies used to lure victims, reflected in higher than expected use of the kindness principle in 2021 and 2022.

***Index Terms*—cybercrime, cryptocurrency, Bitcoin, investment scams, machine learning, natural language processing**

I. Introduction

Fake investment schemes leverage the complexity and abundance of legitimate investment programs. Investment frauds aim to lure victims by offering investment returns that are purported to be higher than average or guaranteed, with very low or no risk attached [1]. Fraudulent schemes also offer investment products that are non-existent [2] or not authorised by regulatory bodies [3]. Research has found fraudsters take advantage of social, economic and cultural factors to manipulate potential investors [4].

For the past decade, cryptocurrencies have risen in value and popularity. According to a consumer research report published in June 2021 by the Financial Conduct Authority (FCA) [5], it was estimated that 42% of the adult population in the United Kingdom were aware of cryptocurrencies in 2019. This percentage increased significantly in 2021 when 78% were aware of them, and more than 2.3 million owned them. This report also found that online advertisements and social media were the primary marketing sources for the interviewed people.

Scammers have taken advantage of this situation. The Australian Competition & Consumer Commission examined the evolution of scams over a period of ten years and identified that most investment scams in 2009 were related to "get rich quick seminars and real estate" whereas in 2019, most of these schemes were connected with cryptocurrencies [6]. This transition has been facilitated by a period of low interest rates from 2009 until 2022 [7], high valuations of cryptocurrencies [8] and the advent of new online platforms for socialising, entertaining and investing.

In more recent times, the COVID-19 pandemic increased the amount of time people spend socialising, shopping and entertaining online [9]. Cybercriminals have taken advantage of this behaviour and of people's fears and needs created by the pandemic. It has been reported that fraudsters have used advertisements on social media and online search engines to lure customers into scams [10]. They have also targeted individuals to lure them into high return investments and other "get rich quick" schemes [11]. The National Cyber Security Centre in the UK blocked or took down more than 5,000

links sent between April and June 2020, involving scams which offered investors high returns in exchange for buying cryptocurrencies including Bitcoin [12].

In this work we primarily focus on advertisements of cryptocurrency-based Ponzi schemes, often referred to as High Yield Investment Programs (HYIPs). These schemes promise very high rates of return (by paying early investors from investments of new ones) [13] or can make use of smart contracts that purport to be unalterable but in reality are Ponzi scams, where fraudsters modify the contract terms and then steal all the investors' money [14]–[16]. These schemes have been linked to a high proportion of investment scams affecting the savings of many investors. According to the *2020 State of Crypto Crime* report from Chainalysis [17], Ponzi schemes accounted for 92% of $8.6B in transactions linked to cryptocurrency-related crime in 2019. The largest program, referred to as "PlusToken scam", was associated to loses of more than $3B affecting more than three million victims. This scheme promised returns between 10% and 30% through advertisements in public groups through WeChat, the most popular messaging app in China, and through its own app where it made use of a referral program that rewarded users for bringing others into the scheme.

Another type of cryptocurrency investment fraud advertisements we delve into includes scams related to Initial Coin Offerings (ICO) [18], Non-Fungible Tokens (NFT) [19], Decentralized Finance (DeFi) [20] and Metamask (a cryptocurrency wallet) [21]. We must clarify that our objective is not to conclude that all projects involving these terms are related to investment scams. Instead, we aim to find those instances where advertisements use these words to lure victims into potential scams offering high levels of return.

Several researchers have studied cryptocurrency-related investment scams [14], [15], [22]–[24]. The first paper inspiring this research investigated Bitcoin-related Ponzi schemes in the online forum Bitcointalk[1] from 2011 to 2016 to analyze features linked to the schemes' success or failure [23]. These features were related with the age of the scammer's account, the level of interaction between scammers and victims, and the involvement of "shills". We build upon this prior work, analysing the evolution of investment scams lures, using the same online forum.

The second paper we use as inspiration is Stajano and Wilson's [25] scam lure typology. Based on several fraudulent cases, they identify seven principles that can be applicable to offline and online scams, namely the authority, dishonesty, distraction, financial, herd, kindness, and time principles. They state that cybercriminals take advantage of the vulnerabilities caused by these precepts and that being aware of these flaws can help design more resilient systems.

We analyze the evolution of advertisements for fraudulent cryptocurrency investment schemes over a period of 12 years (from July 2010 until June 2022). We are interested in understanding how the frequency of these advertisements in Bitcointalk has changed, particularly during the pandemic. We also explore the types of lures used to promote these fraudulent investment schemes.

The objectives and contributions of this paper are to:

1) provide a longitudinal analysis of the evolution of threads and scam-related keywords in the forum over a period of 12 years,
2) design and implement a categorisation criteria to (i) classify threads into overt scams, potential scams, scam comments and not investment scams and (ii) identify the types of actors, lure types and cryptocurrencies used within scam-related threads,
3) evaluate the frequency change of overt and potential scams posted in this forum over time and the lures used to defraud investors.

In §II, we provide a review of related work. In §III, we present our methodology, including the ethical considerations and details of our data collection, annotation and our classifiers. We provide a longitudinal analysis of the data collected and evaluate our results on §IV. We discuss our findings in §V, before concluding (§VI).

II. Related work

There is a substantial body of work analyzing different types of cryptocurrency-based investment scams. Moore et al. [13] were the first to analyze HYIPs. They investigated "aggregator" reputation websites, which monitor these programs and promote them in exchange for a fee. They found no proof of coordinated actions between aggregators and identified that HYIPs that offered lower interest rates and had longer compulsory investment conditions survived for longer periods of time. Subsequently, Drew and Moore [26] used clustering techniques to evaluate the relationships between HYIPs websites. Neisius and Clayton [27] delved into the structural connections between HYIPs and aggregators in more detail, including their revenues and profitability. They identified a key software provider which enabled the existence of more than 50% of HYIPs and aggregators.

Later on, Vasek and Moore [24] were the first to categorise different Bitcoin-related scams. They classified these schemes into Ponzi schemes, mining scams, fake wallets and exchange scams. By using data from Bitcointalk, Cryptohyips and a fraud blocklist from Badbitcoin, they found that the success of a scam relied on large payments from very few victims. Vasek and Moore [23] continued their line of research on Ponzi schemes found on Bitcointalk and highlighted some success features linked to a scam's lifetime. They categorised actors as scammers, victims and shills and stated that actors' prestige was important for a scam lifespan. They also found that scams had a longer life when there was higher shill intervention, but scams had a shorter existence when greater daily communication happened between scammers and victims.

Toyoda et al. [22] also focused on Bitcoin-based HYIPs. Their model analyzed transactions from Bitcoin addresses belonging to HYIPs operators. They built a classifier to predict whether a Bitcoin address belonged to specific HYIP owners.

[1]https://bitcointalk.org

TABLE I: Number of threads/posts with keyword within the post content and thread title

Keyword/phrase	Threads from 3 subforums	Posts from 3 subforums	Threads from all subforums	Posts from all subforums	Ratio of 3 subforums threads/ all subforums threads
Ponzi	3,570	306,779	9,870	4,929,234	36%
HYIP	1,061	166,916	2,610	3,512,873	41%
Rug pull	22	659	916	1,176,457	2%
Doubler	1,014	103,516	3,421	4,310,787	30%
Get rich with Bitcoin/crypto	146	20,512	266	907,384	55%
NFT	34	89,664	4,390	4,238,960	1%
ICO	14	1,034	856	1,710,100	2%
DeFi	0	0	77	436,543	0%
Metamask	91	1,927	1,825	2,147,664	5%

The authors tested their model against a list of 32 HYIP addresses arguing that it accurately detected 93.75% of them correctly.

Other researchers investigated particular scams or types of scams using Bitcoin. Boshmaf et al. [28] used comments on Bitcointalk to gather Bitcoin addresses of users that were a part of the MMM Ponzi scheme. Badawi et al. [29] examined a particular type of Ponzi scheme, the Bitcoin generator scam where websites pretend to generate more Bitcoin after receiving an initial investment. They developed a system that finds new Bitcoin Generator websites using targeted search engine strings.

Ponzi schemes based on other cryptocurrencies such as Ethereum have been researched by Chen et al. [14], [15] and Bartoletti et al. [16]. These authors state to have found smart contracts on the Ethereum blockchain that are in reality Ponzi schemes. They obtain features from the smart contracts code and transaction history and build classifiers to predict whether a Ponzi scheme is disguised as a smart contract.

Other work has been done around cryptocurrency-related scams more broadly. For example, Badawi and Jourdan [30] perform a review of publications about cryptocurrency cyber-attacks, research techniques and data sources used in each study. Their survey includes research of cryptojacking attacks, HYIPs, money laundering, pump and dump schemes and ransomware. Morin et al. [31] analyse plagiarism in cryptocurrency whitepapers and find an increased level of plagiarism in newly introduced ICOs, 19% compared to 4% of actively traded coins. Trozze et al. [32] also provide a systematic literature survey of cryptocurrency-related crime research along with opinions from specialists in this matter. They state that the majority of published investigations have focused on Ponzi schemes, HYIPs and ICO scams. Sapotka et al. [33] use Bitcointalk in order to determine whether a given ICO is fraudulent. Mazzorra et al. [34] use machine learning techniques to spot rug pulls proactively. Kshetri [35] introduces a categorisation criteria for NFT-related scams.

Mackenzie [36] provides an ethnographic account of cryptocurrency trading, claiming that investment scams are indistinguishable from legitimate schemes. However, we note that this is not how Ponzi and pyramid scams are typically advertised. This is accounted for by Mackenzie, who points out that marks may be made to feel complicit and therefore less likely to go to the police to report their involvement. Mackenzie also claims that scam lures create a sense of 'time pressure', which is similar to Stajano & Wilson's [25] time principle.

As we mentioned earlier in §I, some criminals have taken advantage of the social and economic impact that the COVID-19 pandemic has had in people's lives. This has been studied by Bitaab et al. [37] who identify that phishing campaigns increased by 220% between March and April 2020 compared to the same period before the pandemic started. Kemp et al. [38] and Buil-Gil et al. [39] use data from the UK and conclude that cybercrime rose higher than expected due to social behavioral changes caused by the pandemic. Lallie et al. [40] reach a similar conclusion and mention how cybercriminals take advantage of government statements to customize their luring techniques. Jaber and Fritsch [41], also examine how cybercriminals leverage topics around the pandemic to spread malware and phishing links.

Our work is also related to the use of social engineering techniques to deceive investors. Several people have looked into the application of lures in different kinds of illegal activities. Hong et al. [42] evaluate the use of these tactics in "mobile gambling" fraud. Other researchers have explored on the lures used in phishing emails [43]–[45]. Ferreira et al. [43] find that successful phishing campaigns leverage diversion and authoritative techniques. Van der Heijden and Allodi [44], Williams et al. [46] and Quinkert et al. [45] also look into phishing emails and focus on the use of the authoritative principle as the most effective one.

Weber et al. [47] linked the use of social engineering methods with five cryptocurrency-related fraud cases. They use enticement principles [48] found in phishing to review which strategies are effective in cryptocurrency-based scenarios. They describe four examples where fraudsters use the authority principle to persuade victims to share their cryptocurrency wallet security details and to make transfers. To the best of our knowledge, no other study has been done on the types of lures used to attract investors through advertisements of cryptocurrency-based investment scams. This work provides an insight into some of the tactics used by fraudsters to lure victims into these schemes by analyzing conversations at scale on a publicly available online forum.

III. METHODS

A. Data collection

Bitcointalk is the largest online forum focused on cryptocurrencies and was created by Satoshi Nakamoto, who posted the forums' first message in February 2009. As of the date of data collection, August 15, 2022, the forum had more than 3.4M members and a total of 60.7M posts from 1,332,609 topics or threads. We extracted a total of 17.8M posts from the forum from July 2010 until June 2022, from 399,709 threads. This represents 29-30% of all posts/threads in Bitcointalk as of the date of collection. We leveraged the data collectors from the Cambridge Cybercrime Centre[2]. In particular, we used a version of CrimeBot [49] , a crawler specifically created to obtain data from online forums. We must note that our crawler is not able to collect posts that have been deleted. We focused on analysing only those posts written in English which amount to 281,523 threads. We also extracted and examined all posts from three subforums that were investigated by Vasek and Moore [23]: Investor-based games, Games and Rounds, and Scam accusations. The total number of posts from these three subforums is 963,731 which relate to 26,712 threads.

One of our objectives is to show the evolution of scam-related threads and keywords in the forum. For this purpose, we selected two types of keywords and key-phrases to study their frequency appearance over a period of 12 years. The first type of keywords includes words commonly known to be used in advertisements of cryptocurrency-related scams: "Ponzi" (Ponzi scheme) [50], "HYIP" (high yield investment program) [13], "rug pull" or "rugpull" (where developers abandon a cryptocurrency scheme and steal the participants' investments), "doubler" or "doubling" (commonly found on adverts to "Double your Bitcoin") and "get rich with Bitcoin/crypto". We also selected a second type of keywords that are related to cryptocurrencies and are, in theory, related to legitimate projects but which, in some occasions, have been connected to fraudulent transactions: "NFT" (non-fungible token), "ICO" (initial coin offering), "DeFi" (decentralised finance) and "metamask" (a cryptocurrency wallet, which sometimes has been used by criminals in phishing spam campaigns). We used PSQL queries to locate posts that contain these terms within their content or within the thread title. Table I shows the total number of threads and posts with these keywords within them.

We can observe that the number of threads that contain the keywords/phrases "Ponzi", "HYIP", "doubler", and "get rich with Bitcoin/crypto" and belong to the three main subforums represent at least 30% of all the threads in all subforums. However, that is not the case for the term "rug pull", a term mainly used in other parts of the forum.

B. Data annotation

We created a script in R to annotate the data. Our annotation criteria are described in Table II. We extracted a random sample of 130 threads from the three subforums mentioned above and limited the number of responses to a maximum of 10 per thread. This sample contained 1,718 posts. For comparison purposes, we extracted a second random sample of 1,400 threads from the forum at large and selected only the original post without any replies. We annotated both samples based on the same criteria II and used it to train all the models listed in §III-C. After reviewing false positives and false negatives from the predictions obtained, we decided to collect and annotate an additional random sample from the three subforums previously indicated. This sample had 1,100 threads with only the original post and no replies.

After meeting to moderate our annotations of 4,218 posts from 2,630 threads, a Cohen's κ coefficient of 0.908 was obtained between the two reviewers, which reflects almost perfect' agreement according to the criteria by Landis and Koch [51].

C. Classification

We built classifiers to categorise all threads/posts using the same criteria mentioned in §III-B. We used heuristics to identify the types of cryptocurrencies used within the scam-related posts. We compared the performance of four statistical models: Support Vector Machines (SVM) [52], [53], Multinomial Logistic Regression [54], Random Forests [55], and XGBoost [56]. The features included in all models were the post content, thread title, subforum name, and author name.

To pre-process the data, we changed all text to lower case and eliminated any blank inputs and stop-words. We tokenised all input text, and performed word lemmatisation using the NLTK library.[3] We then used the Term Frequency-Inverse Document Frequency (TF-IDF) words weighting [57] to obtain the vector of lexical features.

We split the input data for training and testing using a ratio of 67/33 correspondingly. The training data was unbalanced, which was expected because the majority of posts in the forum are not scam related. To deal with the skewed data distribution, we oversampled the training data using SMOTE [58]. To avoid overfitting the training data, we tuned the models' hyperparameters and used ten-fold crossvalidation.

To evaluate all models, we used the area under the receiver operating characteristic (ROC) curve (AUC) which can be used to measure and compare different classification models' performance [59].

D. Ethical considerations

The department's ethics committee at the University of Cambridge approved this research, which uses data extracted from a publicly accessible forum. The forum's terms of service do not explicitly forbid scraping. The forum has a section that describes current practices about privacy concerns which allows users to participate in the forum without submitting any personal information. The information posted in the forum does not seem to be private information that inadvertently has become publicly available. The forum provides privacy advice

[2]https://www.cambridgecybercrime.uk

[3]http://www.nltk.org

TABLE II: Annotation criteria

Category	Description	Anonymised example
Thread type		
Overt scam	The thread invites others to invest in a scheme explicitly recognised as a scam (Ponzi scheme, HYIP, etc.). The thread title usually has the name/details of the scheme.	Ponzi [url] Please post all of your txids, to and from in this game. The first 50 players will receive and extra 25% After this it will return to the regular 120% Send BTCtc to [address]
Potential scam	The thread invites others to invest in a scheme promising investment returns that are unusually high and/or guaranteed but it does not make specific reference to a Ponzi scheme or a HYIP. We include in this category advertisements for ICOs, cryptocurrency exchanges, mining companies, raffles and gambling adverts only if they offer high rates of return. We do not consider cryptocurrency mixers as investment scams.	Hello, we just released smart platform for bitcoin investors. The idea is really simple: -Invest BTC -We trade it using our softwares (semi-automated trades) -In 7 days you get your BTC quantity doubled. We have several automated bot-softwares, which are tracing altcoins and this way we make multiple profitable trades every day. Escaping dealing with security and your money are safe - no registration, fast, easy and simple.
Scam comment	Relates to investment scams, but is not an invitation to invest. May include people asking for advice for setting up scam-related investments, sharing advice on how to spot a scam, reporting a fraudulent or fake investment scheme (known or unknown as a Ponzi scheme) etc.	We all know that 3.5% interest an hour is impossible and that it's either a Ponzi or a scam. He claims that by investing in mining equipment he can mine 168% of your investment in 48 hours. Which is impossible.
Not scam related	The post content or the thread title is not related to investment scams.	
Scam actor type		
Scam owner	The user invites others to invest in an overt or potential scam. This is usually the first user who starts a thread.	Automatic payments. Send BTC. Get 110% back when the next person sends.
Scam shill	This refers to users that make comments (usually positive) that seem to legitimise a scheme.	Have Anyone Used [url]? to double your bitcoin in 3 hours? here is website [url] Thanks
Scam participant	The user responds to the scam owner knowing that they are participating in a fraudulent scheme.	i have invested here almost 1 week ago and it pays me. just want to know if its still paying or already scam ty
Scam victim	The user claims to have been defrauded or lost from a previous investment.	i lost over 0.095btc ,,,big dreams make big disaster
Scam reporter	The user reports other users claiming that their posts/advertisements are scams. They do not need to have invested and lost themselves.	What happened: Ponzi scheme 250% PROFIT PER 25 DAYS Scammer Profile Link: [url] Website: [url]
Scam commenter	The user discusses investment scams but does not fall into any of the above categories.	Why do people invest in ponzi?
Not scam related	The post does not include anything related to investment scams.	
Scam lure type (adapted from Stajano & Wilson [25])		
Authority principle	The scammer invokes authority such as by demonstrating technical knowledge (e.g. using encryption) or referring to trusted third parties (e.g. Companies House, CloudFlare) to convince users to do things that they would not do otherwise.	UK Registered Company. [Company] is official registered in United Kingdom, you can check its number on companies house here. [url] is being kept on a dedicated server with support and protection from malicious DDoS attacks by hackers.
Dishonesty principle	The scammer encourages others to participate by making them aware that their profit comes from the losses of others.	Your profits will at least double within two days, but do not forget that the project is an investment game and your profits are made from ensuing investors.
Distraction principle	The scammer offers an investment opportunity and provides a lot of irrelevant details.	Here you can earn Ethereum, growing four kinds of vegetables. One acre of garden field gives one vegetable per day. The more acres you have, the more vegetables they give. Attention! Market value of each vegetable will be different. Less grown vegetables will be more expensive.
Financial principle	The scammer takes advantage of users' 'need and greed' to promise enticing options and convince users to make an investment.	SEND BTC TO [address] GET BACK 200% OF YOUR DEPOSIT AT YOUR WALLET ADDRESS WITHIN 48 HOURS.
Herd principle	The scammer refers to the popularity of the scheme to convince victims to not be left out of the investment rewards.	Until now we have more than 7300 Investment, Become now a potential investor.
Kindness principle	The scammer relies on users' willingness to help in order to steal their money.	A community of people providing each other financial help on the principle of gratuitousness, reciprocity and benevolence.
Time principle	The scammer puts time pressure on users so they make rushed and less reasonable choices.	We made a lot of promotion this week for [scheme], tonight at 9PM come and get your 130% payout instantly!
Not scam related	The post does not include anything related to investment scams (although it may be a post on a scam thread).	
Not applicable		

to users, indicating they are aware postings are public. This work focuses on understanding aggregate information and collective behaviour. We do not investigate specific individuals, or attempt to identify forum users. Therefore, this work falls outside the requirement of informed consent, under the British Society of Criminology's Statement of Ethics [60].

IV. Results

A. Longitudinal analysis

Figure 1 shows all threads from all subforums from July 2010 to June 2022 along with the logarithm of Bitcoin's price (transformation used to decrease observed volatility). This shows a pattern of 'bursty' activity, with periods of intense discussion interspersed with quieter periods. In particular, we can see that 2018 was the year with the largest number of threads posted in the forum (770 per day) which followed a sharp decline of more than 80% in 2019 and 2020. The number of daily threads increased slightly in 2021 and have remained constant since then.

Changes in the number of threads seem to coincide with movements in cryptocurrency prices during some time intervals. According to Bitcoin historical prices [61], Bitcoin's price had an annual increase of 5,690.96% between January 1st, 2013 and January 1st, 2014. We observe that the number of daily threads increased more than 625% around the same time. We also see that daily threads reached their maximum between January 1st, 2017 and January 1st, 2018 when Bitcoin's price had a year over year change of 1,244.35%. On the contrary, Bitcoin's price decreased significantly between January 1st, 2014 and January 1st, 2015 with a year over year change of -59.25% and again between January 1st, 2018 and January 1, 2019 when it had an annual change of -71.15%. Similarly, the number of threads decreased significantly around the same time periods.

Figure 2 presents the top 10 subforums from July 2010 and June 2022. We can observe which subforums have become more prevalent over time. For example, in 2013 the most popular subforums were "Beginners & Help" and "Altcoin Discussion" which contained more than 40% and 24% of the total number of threads respectively. This changed in 2014 and 2015 when threads had a more uniform distribution across the top 10 subforums. In 2016 the suforums "Press" and "Digital goods" where the most prominent subforums whereas in 2017 and 2018 "Altcoin discussion" had the largest number of threads. The highest proportion of threads in 2019 and 2020 appear in "Scam accusations".

Figure 3 shows the trend of posts that contain the keywords "Ponzi", "HYIP", "rug pull" (or rugpull) and "Get rich with Bitcoin/crypto" from 2010 to 2022. All graphs show the frequency change in the number of posts that contain the keyword in the thread title or the post content. For this part of our analysis, we focus on the number of posts (instead of the number of threads) because these keywords can be found in several posts within a single thread. We notice that the total number of posts for all keywords peaked in 2014. We can also see that posts with the keywords "Ponzi" and "HYIP" became

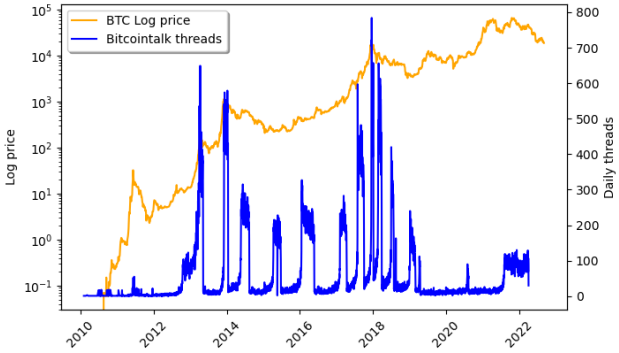

Fig. 1: All threads from July 2010 to June 2022

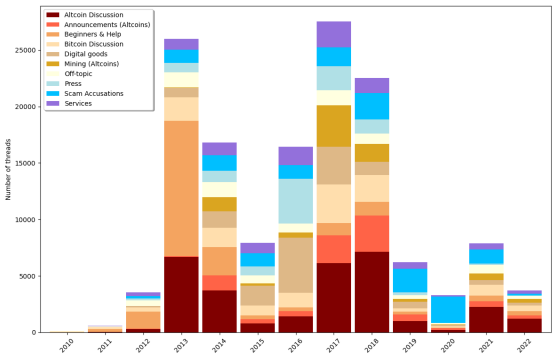

Fig. 2: Top 10 subforums from July 2010 to June 2022

prevalent again towards the end of 2017 and started decreasing in 2018. In the second half of 2021 "HYIP" again became a popular keyword found in posts.

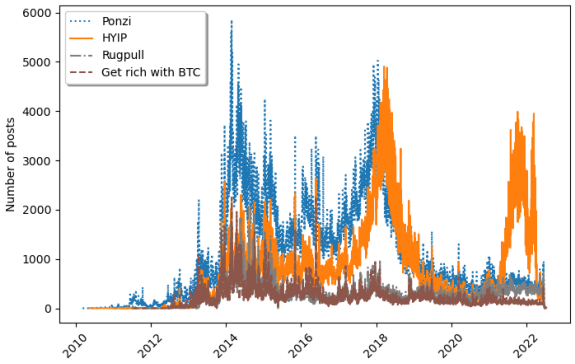

Fig. 3: Posts that contain keywords

B. Predictions by thread type

The XGBoost model showed the best performance (AUC ROC score of 0.953) and less overfitting (when compared to the other three models). We used this model to classify the first post (the 'OP') for each of the 281,523 English-language

threads on Bitcointalk. The classifier predicted that 259,455 threads (92.16%) are not scam related, 16,555 threads (5.88%) relate to potential scams, 323 threads (0.11%) are overt scams and 5,190 threads (1.84%) are scam comments. These numbers are consistent with our annotation process and the data used to train the statistical model.

Figure 4 presents the number of threads categorized as overt scams and potential scams from 2010 to 2022 (we use two separate scales for ease of comparison due to the different magnitudes in the number of threads for each category). This Figure shows that overt scams, which are threads that advertise Ponzi schemes and HYIPs in a specific and open manner, peaked at 14 in one day in 2015. On the other hand, potential scams, which offer exorbitant rates of return in a very short period of time but do not explicitly say that they are Ponzi schemes or HYIPs, appeared most frequently at the beginning of 2018, when 79 of them were found in one day. These advertisements then declined significantly from 2019 until 2021 and have risen slightly in 2022. The maximum number of potential scams also coincides with Bitcoin's price increase.

Figure 5 shows the top 10 subforums where advertisements of potential scams were most frequently located based on the classification of threads over the 12 years of observation. We observe that the subforum "Announcements (Altcoins)" was the most prominent in 2014 and from 2017 until 2019. In 2015 and 2016, the subforum "Investor-based games" contained the majority of threads. More recently, in 2021 and 2022, the subforum "Tokens (Altcoins)" was the most prevalent. We also found "Investor-based games" and "Games and rounds" are the subforums where overt scams are most frequently found. This is consistent with the subforums used by Vasek and Moore [23].

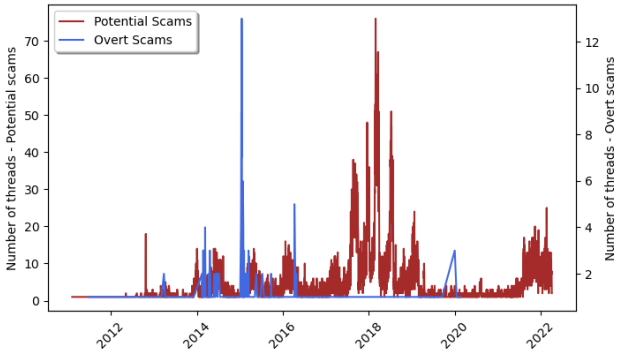

Fig. 4: Overt Scams and Potential Scams

Figure 6 shows the ten most prominent subforums where scam comments are posted. We can see that, according to our classifier, the "Scam accusations" subforum is where the majority of these threads are found over time. We can notice the rising trend of scam accusations from the forum's inception until 2018, when they reached their peak. These subsided slightly in 2019, remained almost constant in 2020, decreased by more than 50% in 2021, and decreased again in 2022.

One of our objectives is to study the impact of the COVID-19 pandemic in the frequency of investment scam advertisements. Figure 7 shows the number of threads classified as of potential scams and scam comments, divided by the total number of threads in the forum (used as a control variable) from September 2019 until June 2022.

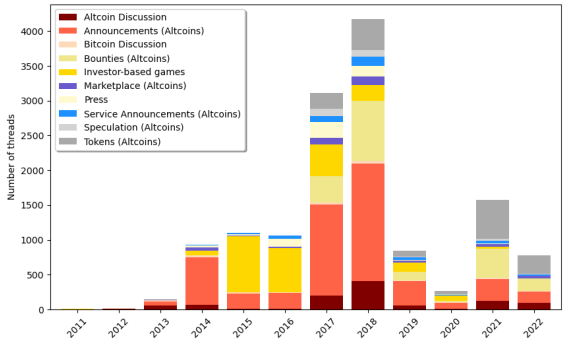

Fig. 5: Potential Scams - Top 10 subforums

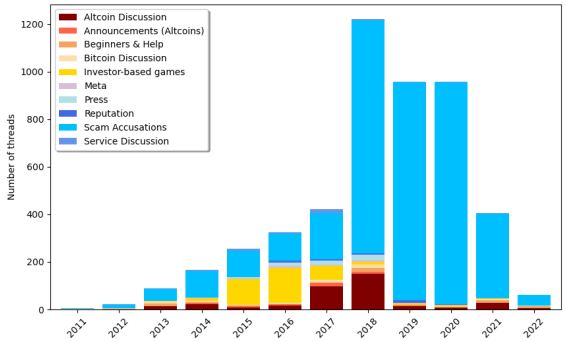

Fig. 6: Scam comments - Top 10 subforums

C. Predictions by scam actor type and lure type

The same statistical models mentioned in §III were used to classify the 16,878 OPs identified as overt and potential scams by scam author type and scam lure type. The XGBoost model again showed the best performance for both classifiers (AUC ROC score of 0.749 and 0.768 respectively). Our training sample (1,313 posts) contained 626 scam owners (47.68%), 396 scam commenters (30.16%), 145 scam reporters (11.04%), 85 scam participants (6.47%), 53 scam shills (4.04%) and 8 scam victims (0.61%). The prediction results for scam actor type were poor since the majority of predictions were attributed to scam owners.

Table III shows the number of scam lure predictions using our typology inspired by Stajano & Wilson's [25]. We use single-label after testing both this and multi-label classification approaches and finding the single-label to be better performing. We can see that the two main types of lures used are the financial principle followed by the distraction and authority

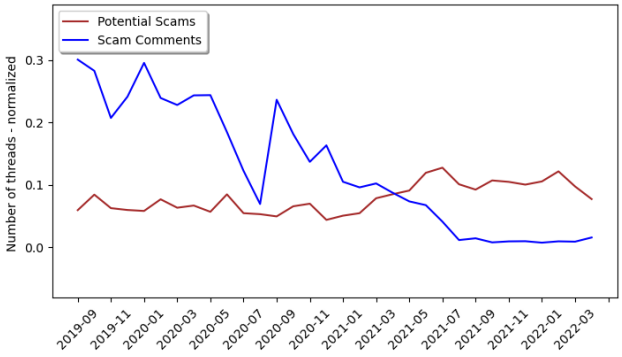

Fig. 7: Potential Scams and Scam Comments from September 2019 to June 2022

principles. We also notice that the time principle is not a tactic used as frequently as expected.

TABLE III: Number of scam-related threads classified by scam lure type

Scam lure type	Number of threads	Percentage
Financial principle	8,218	48.69%
Distraction principle	5,526	32.74%
Authority principle	1,753	10.39%
Herd principle	392	2.32%
Dishonesty principle	353	2.09%
Time principle	349	2.07%
Kindness principle	287	1.70%
Total	16,878	100.00%

Figure 8 displays the mosaic plot of scam lure types across the 12 years of observation. In this plot, tiles that are colored as red/blue have a constraint that is larger/smaller than 2 positive/negative standard deviations from the expected value under the independence hypothesis. The chi-square test confirms that the relationship between years and lures is significant ($\chi^2(66, N = 16,878) = 1,028.69, p < .001$). This shows that in 2015 to 2017, the financial principle was significantly more likely to be found in scam lures. However, from 2018 the financial principle was used less often, and instead the distraction and authority principles were used more frequently than expected. We also see the kindness principle, although used infrequently overall, is used more during 2021 and 2022 (and also 2014).

D. Analysis of connections between predictions and scam-related keywords

We can see in Figure 9 the mosaic plot of the relationship between the scam lure types and the keywords mentioned in §III-A. The corresponding chi-square test shows that the relationship between keywords and lures is significant ($\chi^2(36, N = 3,685) = 793.87, p < .001$). We can observe that the scam-related posts that contain the keyword "Ponzi", are more likely to be found in posts that use the financial, dishonesty and authority principles as a lure. Furthermore, scam advertisements that use the distraction principle are more

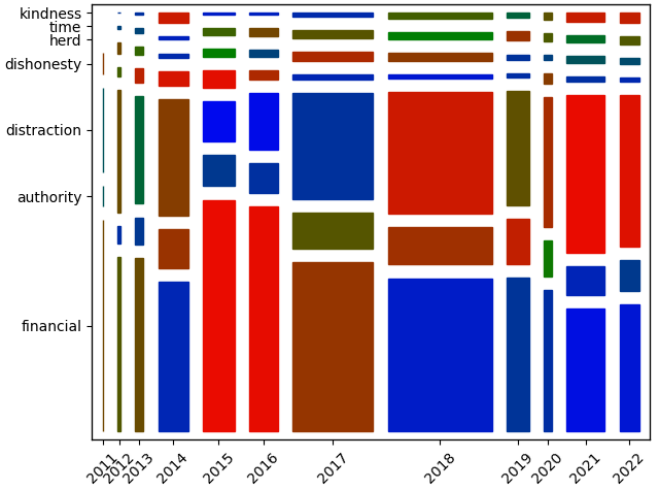

Fig. 8: Mosaic plot of scam lures over time

likely to contain the keywords "NFT", "HYIP", "ICO" and "Metamask" more often.

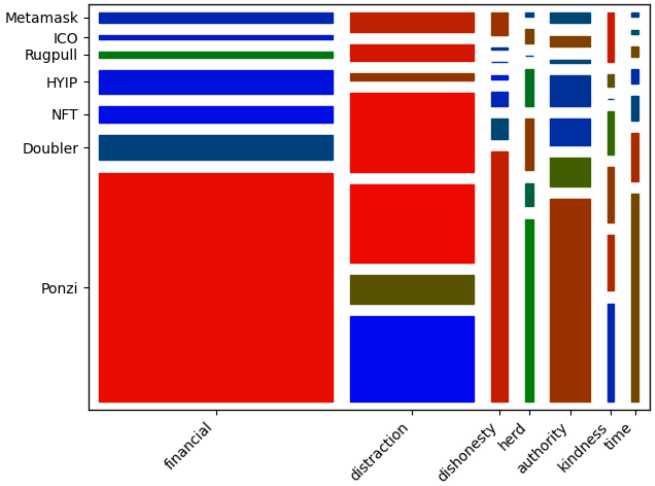

Fig. 9: Scam lure types and keywords

Our modelling also shows that posts classified as overt scams contain the keyword "Ponzi" more than 95% of the time. This is not the case for potential scams, as Figure 10 shows. We notice that "Ponzi" was the keyword most commonly linked to potential scams from 2011 until 2017, 2019 and 2020. However, in 2018 "HYIP" is the most frequent keyword related to these types of posts. In 2021 and 2022, "NFT" seems to take that position.

E. Cryptocurrencies used in investment scams advertisements

We used heuristics to determine the cryptocurrency used in investment scams adverts. The results are shown in Table IV. Bitcoin is the cryptocurrency most frequently used in overt scams, which also use Ethereum, Litecoin and Dogecoin. Potential scams utilize Ethereum followed by Bitcoin, Litecoin and Tether. The majority of threads, however, do not mention

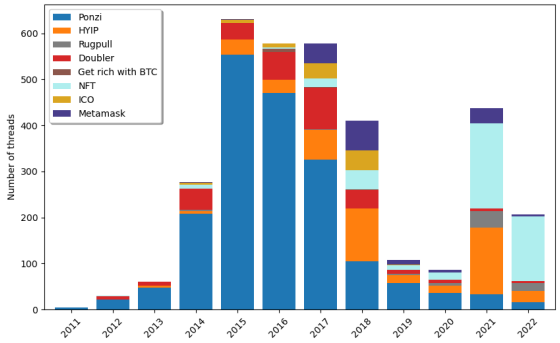

Fig. 10: Potential scams and Keywords

any specific cryptocurrency or advert other new cryptocurrencies, especially in ICO advertisements. While many ICOs use Ethereum to crowdfund, which would intensify our results on potential scams, we hesitate to classify these without explicit mention.

V. Discussion

Our longitudinal analysis presents the evolution of the forum over a 12 year period. Overall, our findings suggest that the popularity of Bitcointalk has waned over time. The changes in popular subforums across the initial years, as shown in Figure 2, such as "Beginners & Help" and "Altcoin Discussion" in 2013, to a more uniform distribution in 2014 and 2015, could be related to moderation in the forum. As we can see in Figure 1, the number of daily threads in the forum increased exponentially in 2013, starting at less than 100 threads per day at the beginning of that year, and finishing at 625 per day towards the end of the same year. We observe a similar trend in 2014 followed by relative decrements from 2015 until the end of 2017. Subsequently, the forum's threads reached their maximum in 2018 and have decreased by more than 80% since then. It is possible that as cryptocurrencies have become more mainstream, and the relevant marketing channels have become more diverse, discussions (and advertisements) have moved to other online platforms. Exploring scam lures on other sites will be the subject of future work. Another explanation for the forum's decline is provided by information asymmetry [62], which suggests that genuine traders who cannot differentiate themselves from fraudsters will leave the market.

We also notice that in 2019 and 2020, the "Scam Accusations" subforum accounted for the largest proportion of threads in the forum. Topics in this specific subforum have changed over the years. Discussions found in 2012 and 2013 refer usually to personal complaints of users against other participants in the forum. These accusations usually warn participants about lack of payments or responses from certain individuals. During this time, it was harder to purchase cryptocurrencies via cryptocurrency exchanges, so many users bought and sold Bitcoin directly from each other, forcing the issue of user reputation. A smaller number of topics in these early years are around Ponzi schemes and cryptocurrency exchanges such as Mt. Gox. We observe that personal accusations have continued throughout the 12 years of observation. However, the proportion of topics related to 'industrialized' scams (such as Ponzi schemes, crypto-mining scams, scam exchanges, etc.) have increased significantly and seem to be the topics with more views and responses in 2021 and 2022. This could be another reason for the decline in the forum's popularity.

We observe volatile frequency changes in the number threads across the 12 years of observation. These changes seem to coincide at some points in time, with movements of Bitcoin's prices. Larger numbers of threads are observed when Bitcoin's price increases and sharp declines are seen when there is a fall in the cryptocurrency valuation. These patterns are consistent with those observed by other researchers who have analyzed the links between cryptocurrency price "bubbles" and the quantity of conversations that appear to exacerbate them [63]–[65].

Our thread type classifier was applied to all the 281,523 OPs extracted from the forum. We identified 16,878 overt and potentials scams advertisements. These threads were most frequently located in four subforums overtime, namely "Investor-based games", "Games and rounds" (consistent with the subforums used by Vasek and Moore [23]), "Announcements (Altcoins)" and "Tokens (Altcoins)". As shown in Figure 6, the majority of threads classified as scam comments (5,190) were found in the "Scam accusations" subforum. This reflects the accuracy of our thread type classifier and its potential application to other datasets with similar lexicographic styles.

Figure 3 indicates that posts with scam-related keywords such as "Ponzi", "HYIP", "rug pull" and "get rich with Bitcoin/crypto" reached their first peak in 2014. Subsequently, in 2018 "Ponzi" and "HYIP" became widespread again and we can observe a transition from the former to the latter. The incidence of these keywords diminished since then, with the exception of posts including "HYIP" within them. These posts re-emerged again in 2021 possibly due to the effects of the COVID-19 pandemic.

To analyze changes in the occurrence of investment scam advertisements during the pandemic (Figure 7) we also look at the ratio of potential scams threads to the total number of threads in the forum (used as a control variable). We observe that potential scams increased by 24.1% from March 2020 until March 2021 and 24.2% between March 2021 and March 2022. This is consistent with some reports [10], [66] mentioning the growth of cryptocurrency-related crime during the pandemic. We applied the same control variable to scam comments and identify that these decreased by 55.2% between March 2020 and March 2021, and decreased again by 91.5% from March 2021 until March 2022. This decrease could be justified by the diminishing popularity of Bitcointalk.

We classified all scam-related posts by lure type and found the financial principle to be the lure tactic that was used more frequently than expected between 2015 and 2017 (as shown in Figure 8). We can detect a transition to the distraction principle in 2018 and between 2019 and 2022. The authority

TABLE IV: Number of scam-related threads and their cryptocurrency

Cryptocurrency	Number of threads	Percentage	Potential Scams	Overt Scams
Bitcoin	1,520	9.0%	1,470	50
Ethereum	5,231	31.0%	5,200	31
Tether	443	2.6%	443	0
Cardano	104	0.6%	104	0
Litecoin	720	4.3%	711	9
Dogecoin	393	2.3%	386	7
Shiba Inu	49	0.3%	49	0
None mentioned or other	8,418	49.9%	8,530	0
Total	16,878	100%	16,781	97

principle appeared more often than expected in 2014 and 2019. These strategies explain the lures used to defraud investors over time. People that participate knowingly and willingly in these schemes are aware of the risks and potential rewards involved in Ponzi schemes and HYIPs. However, there are other individuals that look for investment opportunities that cannot be found in 'traditional' venues. It is important that investors are aware of schemes that are "too good to be true" and that seem attractive due to the lures used by criminals to entice their victims.

We can see that the kindness principle is found more frequently than expected during the pandemic, specifically in 2021 and 2022. It is possible that fraudsters took advantage of the effects of lockdowns on people's lives and their lack of social contact.

We also realise that the time principle is not a tactic used as frequently as expected. This challenges the findings by Mackenzie [36] but could be linked to the pyramid structure of Ponzi schemes and HYIPs, which usually pay earlier investors with late investors' money. In these schemes, fraudsters probably would encourage people to invest as soon as possible but they would not use the tactic of mentioning that the project is about to close in order to do so.

VI. Conclusion

The objective of this paper was to analyze the evolution of investment scam adverts, and the types of lures and keywords related to them, identified in Bitcointalk from July 2010 until June 2022. To the best of our knowledge, no other work has previously used such a large dataset to examine the types of lures used to attract investors through advertisements of cryptocurrency-based investment scams.

Our longitudinal analysis shows that the number of threads in the forum increased in an exponential manner in 2013 and 2014, reached its peak at the beginning of 2018 and has decreased more than 80% since then. After collecting more than 29% of all posts in the forum, we evaluated the frequency changes in the number of posts with specific keywords linked to investment scams. We observe a transition from "Ponzi", which was the keyword most frequently mentioned in 2014 and 2018, to "HYIP", most commonly found in 2018 and 2021.

We designed and implemented a classification criteria to categorise posts and identify overt and potential scams, scam comments and not investment scam related posts. We annotated 4,218 posts from 2,630 threads to train four machine learning statistical models. We used these models to classify all English-language OPs by thread type and identified that overt scams were most frequently advertised in 2015. Furthermore, potential scams reached their peak in 2018, subsequently declined and increased again in 2021 during the COVID-19 pandemic. Therefore, we observe a positive influence of the pandemic in the number potential scams identified by our classifier but a negative impact in the number of scam comments found.

We used heuristics to find the types of cryptocurrencies used within scam advertisements and found that Ethereum is the cryptocurrency most commonly found in potential scams. On the other hand, overt scams included Bitcoin most frequently within their post content or thread title.

We identified the types of scam actors are behind the scam-related posts and classified overt and potential scams by lure type, using a customized typology based on the principles identified by Stajano and Wilson [25]. We found the financial principle to be the most popular lure from 2015 until 2017. This principle was used more often than expected while "Ponzi" was the keyword most commonly found around the same years of observation. We identified a transition into the authority and distraction principles from 2018 until 2022. The use of these principles coincide with the periods when "HYIP" was identified most frequently within posts.

We realised that the time principle is not a tactic used as frequently as expected. On the contrary, the kindness principle has been leveraged by fraudsters during the pandemic years of 2021 and 2022.

Our work provides an insight into some of the tactics used by criminals to lure victims into investment scams by analyzing conversations at scale on a publicly available online forum. Being aware of the techniques used by cybercriminals can help design more robust systems against these types of fraudulent schemes.

Finally, we plan to use our thread type and scam lure type classifiers in future work to explore investment scams advertisements in other platforms. Given the existing work that has shown that how cryptocurrencies are discussed differs quite broadly across platforms based on, among other factors the size and interconnectedness of the community [67], we would be interested to see these effects here.

VII. Acknowledgments

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949127) (for GAS and AH), and the National Science Foundation [grant number CNS-1849729] (for MV).

References

[1] Federal Bureau of Investigation, "Business and investment fraud," 2022. [Online]. Available: https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/business-and-investment-fraud

[2] Financial Conduct Authority, "Cryptoasset investment scams," 2022. [Online]. Available: https://www.fca.org.uk/scamsmart/cryptoasset-investment-scams

[3] ——, "Unauthorised firms and individuals," 2022. [Online]. Available: https://www.fca.org.uk/consumers/unauthorised-firms-individuals

[4] I. M. Chiluwa and I. Chiluwa, ""we are a mutual fund:" how ponzi scheme operators in nigeria apply indexical markers to shield deception and fraud on their websites," *Social semiotics*, pp. 1–26, 2020.

[5] M. Karim and G. Tomova, "Research note: Cryptoasset consumer research 2021," 2021. [Online]. Available: https://www.fca.org.uk/publications/research/research-note-cryptoasset-consumer-research-2021#lf-chapter-id-overview-of-key-findings

[6] Australian Competition & Consumer Commission, "Targeting scams 2019, a review of scam activity since 2009," 2019. [Online]. Available: https://www.accc.gov.au/publications/targeting-scams-report-on-scam-activity/targeting-scams-2019-a-review-of-scam-activity-since-2009

[7] O. Data, "Long-term interest rates," 2022. [Online]. Available: https://data.oecd.org/interest/long-term-interest-rates.htm#indicator-chart

[8] Coinmarketcap, "Total cryptocurrency market cap," 2022. [Online]. Available: https://coinmarketcap.com/charts/

[9] United Nations Conference on Trade and Development, "Covid-19 has changed online shopping forever, survey shows," 2020. [Online]. Available: https://unctad.org/news/covid-19-has-changed-online-shopping-forever-survey-shows

[10] Financial Times, "Fraudsters use pandemic fears to part victims from their cash," 2020. [Online]. Available: https://www.ft.com/content/3aaa9447-de89-4b13-9714-fc909a1209cd

[11] National Crime Agency, "Beware fraud and scams during covid-19 pandemic fraud," 2020. [Online]. Available: https://nationalcrimeagency.gov.uk/news/fraud-scams-covid19

[12] National Cyber Security Centre, "Thanks a million: British public help reach major milestone in fight against scammers," 2020. [Online]. Available: https://www.ncsc.gov.uk/news/british-public-help-reach-milestone-against-scammers

[13] T. Moore, J. Han, and R. Clayton, "The postmodern Ponzi scheme: Empirical analysis of high-yield investment programs," in *Financial Cryptography and Data Security*, ser. Lecture Notes in Computer Science, vol. 7397. Springer, 2012, pp. 41–56. [Online]. Available: https://dx.doi.org/10.1007/978-3-642-32946-3_4

[14] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, and Y. Zhou, "Detecting ponzi schemes on ethereum: Towards healthier blockchain technology," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 1409–1418. [Online]. Available: https://doi.org/10.1145/3178876.3186046

[15] W. Chen, Z. Zheng, E. C. . Ngai, P. Zheng, and Y. Zhou, "Exploiting blockchain data to detect smart ponzi schemes on ethereum," *IEEE Access*, vol. 7, pp. 37 575–37 586, 2019.

[16] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, "Dissecting ponzi schemes on ethereum: identification, analysis, and impact," *Future Generation Computer Systems*, vol. 102, pp. 259–277, 2020.

[17] Chainalysis, "The 2020 state of crypto crime," Chainalysis, Tech. Rep., 2020.

[18] Merriam-Webster.com Dictionary, "Initial coin offering," 2022. [Online]. Available: https://www.merriam-webster.com/dictionary/initial%20coin%20offering

[19] Britannica, "Non-fungible token," 2022. [Online]. Available: https://www.britannica.com/topic/non-fungible-token

[20] Ethereum.org, "Decentralized finance," 2022. [Online]. Available: https://ethereum.org/en/defi/

[21] Metamask, "Metamask," 2022. [Online]. Available: https://metamask.io

[22] K. Toyoda, P. Takis Mathiopoulos, and T. Ohtsuki, "A novel methodology for HYIP operators' bitcoin addresses identification," *IEEE Access*, vol. 7, pp. 74 835–74 848, 2019.

[23] M. Vasek and T. Moore, "Analyzing the Bitcoin Ponzi scheme ecosystem," in *Fifth Workshop on Bitcoin and Blockchain Research*, ser. Lecture Notes in Computer Science. Springer, 2018.

[24] ——, "There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams," ser. Lecture Notes in Computer Science, vol. 8975. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 44–61.

[25] F. Stajano and P. Wilson, "Understanding scam victims: seven principles for systems security," *Communications of the ACM*, vol. 54, no. 3, pp. 70–75, 2011.

[26] J. Drew and T. Moore, "Automatic identification of replicated criminal websites using combined clustering," in *2014 IEEE Security and Privacy Workshops*, 2014, pp. 116–123.

[27] J. Neisius and R. Clayton, "Orchestrated crime: The high yield investment fraud ecosystem," in *2014 APWG Symposium on Electronic Crime Research (eCrime)*, vol. 2014-. IEEE, 2014, pp. 48–58.

[28] Y. Boshmaf, C. Elvitigala, H. Al Jawaheri, P. Wijesekera, and M. Al Sabah, "Investigating MMM P†onzi scheme on bitcoin," in *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, 2020, pp. 519–530.

[29] E. Badawi, G.-V. Jourdan, G. Bochmann, and I.-V. Onut, "An automatic detection and analysis of the bitcoin generator scam," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020, pp. 407–416.

[30] E. Badawi and G.-V. Jourdan, "Cryptocurrencies emerging threats and defensive mechanisms: A systematic literature review," *IEEE Access*, vol. 8, pp. 200 021–200 037, 2020.

[31] A. Morin, M. Vasek, and T. Moore, "Detecting text reuse in cryptocurrency whitepapers," in *IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2021, Sydney, Australia, May 3-6, 2021*. IEEE, 2021, pp. 1–5.

[32] T. Arianna, J. Kamps, E. A. Akartuna, F. J. Hetzel, K. Bennett, T. Davies, and S. D. Johnson, "Cryptocurrencies and future financial crime," *Crime Science*, vol. 11, no. 1, 2022.

[33] N. Sapkota, K. Grobys, and J. Dufitinema, "How much are we willing to lose in cyberspace? on the tail risk of scam in the market for initial coin offerings," 2020. [Online]. Available: https://ssrn.com/abstract=3732747

[34] B. Mazorra, V. Adan, and V. Daza, "Do not rug on me: Leveraging machine learning techniques for automated scam detection," *Mathematics*, vol. 10, no. 6, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/6/949

[35] N. Kshetri, "Scams, frauds, and crimes in the nonfungible token market," *Computer*, vol. 55, no. 4, pp. 60–64, 2022.

[36] S. Mackenzie, "Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial," *British Journal Of Criminology*, 2022. [Online]. Available: https://doi.org/10.1093%2Fbjc%2Fazab118

[37] M. Bitaab, H. Cho, A. Oest, P. Zhang, Z. Sun, R. Pourmohamad, D. Kim, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, "Scam pandemic: How attackers exploit public fear through phishing," in *2020 APWG Symposium on Electronic Crime Research (eCrime)*, 2020, pp. 1–10.

[38] S. Kemp, D. Buil-Gil, A. Moneva, F. Miró-Llinares, and N. Díaz-Castaño, "Empty streets, busy internet. a time series analysis of cybercrime and fraud trends during covid-19," *Journal of Contemporary Criminal Justice*, Jul. 2021.

[39] D. Buil-Gil, F. Miró-Llinares, A. Moneva, S. Kemp, and N. Díaz-Castaño, "Cybercrime and shifts in opportunities during covid-19: a preliminary analysis in the uk," *European Societies*, vol. 23, no. sup1, pp. S47–S59, 2021. [Online]. Available: https://doi.org/10.1080/14616696.2020.1804973

[40] H. S. Lallie, L. A. Shepherd, J. R. Nurse, A. Erola, G. Epiphaniou, C. Maple, and X. Bellekens, "Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," *Computers & Security*, vol. 105, p. 102248, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404821000729

[41] A. N. Jaber and L. Fritsch, "Covid-19 and global increases in cybersecurity attacks: Review of possible adverse artificial intelligence attacks," in *2021 25th International Computer Science and Engineering Conference (ICSEC)*, 2021, pp. 434–442.

[42] G. Hong, Z. Yang, S. Yang, X. Liaoy, X. Du, M. Yang, and H. Duan, "Analyzing ground-truth data of mobile gambling scams," in *2022 IEEE Symposium on Security and Privacy (SP)*, 2022, pp. 2176–2193.

[43] A. Ferreira and G. Lenzini, "An analysis of social engineering principles in effective phishing," in *2015 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, 2015, pp. 9–16.

[44] A. Van Der Heijden and L. Allodi, "Cognitive triaging of phishing attacks," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1309–1326.

[45] F. Quinkert, M. Degeling, and T. Holz, "Spotlight on phishing: A longitudinal study on phishing awareness trainings," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, L. Bilge, L. Cavallaro, G. Pellegrino, and N. Neves, Eds. Cham: Springer International Publishing, 2021, pp. 341–360.

[46] E. J. Williams, J. Hinds, and A. N. Joinson, "Exploring susceptibility to phishing in the workplace," *International Journal of Human-Computer Studies*, vol. 120, pp. 1–13, 2018.

[47] K. Weber, A. E. Schütz, T. Fertig, and N. H. Müller, "Exploiting the human factor: Social engineering attacks on cryptocurrency users," in *Learning and Collaboration Technologies. Human and Technology Ecosystems*, P. Zaphiris and A. Ioannou, Eds. Cham: Springer International Publishing, 2020, pp. 650–668.

[48] A. Ferreira, L. Coventry, and G. Lenzini, "Principles of persuasion in social engineering and their use in phishing," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2015, pp. 36–47.

[49] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "Crimebb: Enabling cybercrime research on underground forums at scale," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. International World Wide Web Conferences Steering Committee, 2018, p. 1845–1854.

[50] T. Frankel, *The Ponzi scheme puzzle: A history and analysis of con artists and victims*. Oxford University Press, 2012.

[51] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, p. 159, 1977.

[52] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ser. COLT '92. New York, NY, USA: Association for Computing Machinery, 1992, p. 144–152. [Online]. Available: https://doi.org/10.1145/130385.130401

[53] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, p. 273–297, Sep. 1995. [Online]. Available: https://doi.org/10.1023/A:1022627411411

[54] D. Jurafsky and J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 02 2008, vol. 2.

[55] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[56] R. Shaw, "Xgboost: A concise technical overview, www.kdnuggets.com/2017/10/xgboost-concise-technical-overview," 2017. [Online]. Available: https://www.kdnuggets.com/2017/10/xgboost-concise-technical-overview.html?cv=1

[57] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[58] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953

[59] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320396001422

[60] "British society of criminology. statement of ethics 2015." 2015. [Online]. Available: https://www.britsoccrim.org/ethics/

[61] I. Webster, "Bitcoin historical prices," 2020. [Online]. Available: https://www.officialdata.org/bitcoin-price

[62] G. A. Akerlof, "The market for 'lemons': quality uncertainty and the market mechanism," *Market Failure or Success*, p. 66, 1970.

[63] D. Garcia and F. Schweitzer, "Social signals and algorithmic trading of bitcoin," *Royal Society open science*, vol. 2, no. 9, p. 150288, 2015.

[64] B. Alvarez-Pereira, M. Ayres, A. M. G. Lopez, S. Gorsky, S. Hayes, Z. Qiao, and J. Santana, "Network and conversation analyses of bitcoin," in *Proc. Complex Syst. Summer School*, 2014.

[65] R. C. Phillips and D. Gorse, "Predicting cryptocurrency price bubbles using social media data and epidemic modelling," in *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2017, pp. 1–7.

[66] Financial Times, "The lawless world of crypto scams," 2022. [Online]. Available: https://www.ft.com/content/5987649e-9345-4eae-a4b8-9bfb0142a2ab?desktop=true&segmentId=7c8f09b9-9b61-4fbb-9430-9208a9e233c8

[67] K. Caliskan, "Data money makers: An ethnographic analysis of a global cryptocurrency community," *The British Journal of Sociology*, vol. 73, no. 1, pp. 168–187, 2022.