University of Kent

Kent Academic Repository

Baruwa, Zsofia, Bhattacherjee, Sanjay, Chandnani, Sahil Rey and Zhu, Zhen (2025) *User Perceptions of Cryptocurrency Attacks – Extended Abstract.* In: Proceedings: 2025 Crypto Valley Conference CVC 2025. Crypto Valley Conference on Blockchain Technology (CVCBT) . pp. 138-153. IEEE Xplore, United States of America ISBN 979-8-3315-8749-9.

Downloaded from
https://kar.kent.ac.uk/110714/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.1109/CVC65719.2025.00021

This document version
Author's Accepted Manuscript

DOI for this version

Licence for this version
UNSPECIFIED

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

User Perceptions of Cryptocurrency Attacks – Extended Abstract

Zsofia Baruwa*
Kent Business School
University of Kent
Canterbury, UK
zb78@kent.ac.uk

Sanjay Bhattacherjee
School of Computing
University of Kent
Canterbury, UK
s.bhattacherjee@kent.ac.uk

Sahil Rey Chandnani
School of Computing
University of Kent
Canterbury, UK
sahilreycc@gmail.com

Zhen Zhu
Kent Business School
University of Kent
Canterbury, UK
z.zhu@kent.ac.uk

***Abstract*—This work is the first study on the perceptions of social media users about cryptocurrency attacks. The double-spending or 51% attack being the most fundamental attack on cryptocurrencies, it is the focus of this study. As a first step, we create a first-of-its-kind comprehensive list of 31 events of 51% attacks on various proof-of-work cryptocurrencies, showing that these events are quite common. This list contradicts the general perception about the security of cryptocurrencies, particularly portrayed in the Executive Order establishing a Strategic Bitcoin Reserve and a Digital Asset Stockpile in the US. We design the methodologies for our new study of user perceptions around these attacks. We create datasets containing tweets from the time of the attack events, and compare them with benchmark data from normal times. We define parameters for profiling these datasets based on user perceptions – sentiments and emotions. We study the variation of these perception profiles, when a cryptocurrency is under attack and the benchmark otherwise, between multiple attack events of the same cryptocurrency, and between different cryptocurrencies. Our results confirm some expected overall behaviour and reactions while providing nuanced insights that may not be obvious or may even be considered surprising. Our code and datasets are publicly accessible.**

***Index Terms*—Blockchain, cryptocurrency, double-spending, 51% attack, sentiment analysis, emotion detection, perception profiling.**

I. Introduction

Bitcoin [1], the first cryptocurrency was deployed in 2009. Since then, there has been a flurry of new cryptocurrencies [2] that are in one way or the other, adaptations of the blockchain protocol underlying Bitcoin with a total market capitalisation of almost 3 trillion US dollars[1]. Despite a general lack of public understanding about how cryptocurrencies work and their potential impacts [3]–[8], this considerable interest of investors in their economic promise is undeniable. The US has recently created a strategic Bitcoin reserve with the understanding that Bitcoin has "never been hacked", and a digital asset stockpile that will include other cryptocurrencies [9]. While the long-term effects of these steps remain to be seen, it is generally being considered as a "pivotal moment in the U.S. government's approach to digital assets, and possibly even in monetary history" [10].

Cryptocurrency (or any financial instrument) has value as long as it has its users' trust. The most fundamental aspect of the users' trust in cryptocurrencies is the security of the underlying technology – the assurance that their value will not be lost due to an attack. *The most important attack on a cryptocurrency is called the 51% attack that defeats the fundamental immutability requirement of its underlying blockchain* [1], [11]–[14]. It allows the attacker to rewrite the transaction history recorded in the blockchain ledger. As a result, transactions denoting transfer of coins are removed from the history, allowing the coins to be spent again. Hence it is also called the *double-spending* attack. Preventing this attack is the primary security goal of a blockchain, as described in Nakamoto's whitepaper [1] introducing Bitcoin and blockchains to the world.

Cryptocurrencies are decentralised and distributed systems. In proof-of-work (PoW) systems like Bitcoin, they are maintained through consensus among entities called *miners*. Miners contribute new blocks to the underlying blockchain[2]. A 51% attack is conducted by (subsets of) malicious miners controlling *more than half* of the total mining resources invested in the cryptocurrency system at that point in time[3] – hence the name[4]. Other attacks on blockchains have emerged eventually as the blockchain ecosystem has grown and evolved, but the 51% attack remains as the fundamental security concern of a blockchain. A brief description of the 51% attack is provided in Appendix B.

There is an interdependence between the security of a cryptocurrency, user perceptions about its security (their trust in the system), and its price [7]. As its price increases reflecting greater trust in the system, attacking the system becomes more difficult. Conversely, if a cryptocurrency system is attacked, its price decreases reflecting loss of trust. Hence

The authors would like to thank the Institute of Cyber Security for Society (iCSS), University of Kent, for the seed corn funding provided to this project. They would also like to thank the anonymous reviewers for their detailed comments that helped to significantly improve the quality of the work.

* Corresponding Author

[1]As on July 20, 2025 according to https://coinmarketcap.com/

[2]In proof-of-stake systems, these entities are the stakers/validators. However, the cryptocurrencies considered in this work are all proof-of-work based systems maintained by miners. So we call these entities miners in general.

[3]This is assuming that the network is fully connected. Double spending attacks can be conducted by controlling a smaller percentage of the network hash-rate and splitting the network [12].

[4]The estimated cost of a 51% attack on various proof-of-work cryptocurrencies is regularly updated at https://www.crypto51.app/.

"community trust" is a key parameter in the security score of blockchain-based systems, as provided by popular services like CertiK Skynet [15][5]. We point out that *user perceptions regarding the security of cryptocurrencies are intermediary between attacks and their effect on the price*. (See Appendix C for further details.) Previous works [16], [17] studying user trust on cryptocurrencies have considered aspects like lack of institutional backing, pseudonymity of users, their legal status, volatility of value, reputation of users, etc. In particular, there is a well known effect of peer-influence via the cryptocurrency market itself on their price [18]. However, the effect that any kind of attack on a cryptocurrency has on users' perceptions has not been studied before. In this work, we concentrate on this unexplored problem – *how do attacks on cryptocurrencies affect users' perceptions of them*?

Being a major social media platform, X (formerly Twitter) has been increasingly influential in generating and portraying perceptions [19]–[21]. These perceptions and the consequent price volatility of financial instruments affect investment decisions in real time [22]–[26]. Twitter data are primarily text-based and therefore are typically analysed with natural language processing (NLP) methods such as *sentiment analysis* and *emotion detection* [24], [25], [27], [28]. Whereas sentiment analysis often classifies texts into three polarity levels (i.e. negative, neutral, or positive), emotion detection is focused on more granular multi-class identification of human emotions such as happiness, sadness, surprise, fear and anger [29]. Both sentiments and emotions detected from Twitter do correlate with financial market movements [30] and such data can be used to predict price movements, days ahead [31].

Given that the 51% attack is the most fundamental security risk for cryptocurrencies [1], and user perceptions are well established as market movement predictors [7], we frame the following questions.

- How common are 51% attacks on proof-of-work cryptocurrencies?
- How much and how quickly does such an attack reflect on social media?
- What sentiments and emotions do the users generally have about these cryptocurrencies and how do they vary at the time of a 51% attack?
- Do users of all cryptocurrencies demonstrate similar sentiments and emotions?

Various other attacks on cryptocurrency related systems are already known to be very common, causing huge losses to users [32]–[36][6]. Given the ubiquity of attacks on cryptocurrencies, a common mechanism for their detection and notification to users is of importance. In fact, users have not been able to effectively defended themselves due to the lack of such notifications [34]. We hope that this work will lead to a unified attack detection and user notification mechanism for cryptocurrencies based on social media activities.

A. Our contributions

This is the first study to profile the effects of any kind of attack on cryptocurrencies as expressed in the sentiments and emotions of social media users. Given the importance of the 51% attack, it is the obvious choice over other kinds of attacks. By examining how users perceive and react to such attacks, our results provide insights into the trust dynamics within digital ecosystems. To begin with, a list of 51% attack events is required for such a study. So our first major contribution is to *manually* create a comprehensive historical timeline of 51% attack events, providing cross-references of online sources for most events. To our knowledge, only one such list containing 14 attack events has been reported earlier [37]. These 14 attacks are on 13 mostly minor coins by their market shares, and just one repeated attack[7]. We have identified a total of 31 events of 51% attacks on 20 different cryptocurrencies, since the introduction of Bitcoin in 2009, up to August 2021[8]. Our list reveals that 6 out of these 20 cryptocurrencies have been attacked multiple times, all of which had significant market shares at the time of the attacks. They are Ethereum Classic (4 times), Bitcoin Gold (3 times), Bitcoin SV (2 times), Litecoin Cash (2 times), Verge (2 times) and Vertcoin (2 times). This is presently the most comprehensive list of such real-life attack events to the best of our knowledge. It shows that 51% *attack events have been quite common in the cryptocurrency ecosystem*, contrary to common belief, as particularly portrayed in the Executive Order to establish a Strategic Bitcoin Reserve and a Digital Asset Stockpile in the US [9][9]. *Our list of* 51% *attack events should raise public awareness while informing governments about fundamental security risks often faced by proof-of-work cryptocurrencies.*

Our timeline of attacks originated from our observations of the cryptocurrency space over the years, especially through Twitter. We explored blockchain and cryptocurrency news portals and discussion fora, conducted general internet search, and searches on Twitter, to finalise the timeline through manual scrutiny. Unlike [37], we have provided manually checked references for all events – at least two cross-checked references in most cases[10]. While our comprehensive manual search for attack events may not be fool-proof and hence our list may not be exhaustive, we also do not see an easy way to automate the creation of an *exhaustive list* with limited resources. Here, we clarify that the best conceivable way to detect a 51% attack (or a threat thereof), is to keep monitoring the underlying

[5]The security score for Bitcoin is available at: https://skynet.certik.com/projects/bitcoin.

[6]The most recent attack on the cryptocurrency exchange Bybit is perhaps the "largest ever single crypto heist in history" leading to a theft of over $1.5 billion. https://thehackernews.com/2025/02/bybit-confirms-record-breaking-146.html

[7] [37] analysed the effect of 51% attacks on cryptocurrency prices, without considering user emotions.

[8]We did not have the resources to continue the manual investigations thereafter.

[9]The reserve currently holds approximately 200,000 BTC, worth about $17.5 billion at today's prices, with plans to expand it not only with Bitcoin but also Ethereum and three other non-PoW cryptocurrencies [38].

[10] [37] did not mention how they created their list; they did not provide references corroborating the events in their list.

blockchain itself for its forks and whether they are due to a 51% attack [39]. However, constructing such detection systems for each of the thousands of cryptocurrencies that exist today would be very costly.

With the timeline at hand, we identified a methodology to trace the events on Twitter. We chose a common period of investigation for the events. This started from the day before the attack to six days after the attack ended. For an attack event E_i on our timeline, we have compared the volume (i.e. number) of all tweets on the cryptocurrency ($|\mathcal{W}_i|$) with the volume of the subset of tweets that are explicitly about the attack ($|\mathcal{A}_i|$). We have also compared these two volumes with that of a benchmark dataset ($|\mathcal{B}_i|$) from a time when the respective cryptocurrency was not under attack. In most cases, the volume increased during the attack in comparison to the benchmark period when there is no attack. At the time of the attack, most of the tweets are related to the attack. To understand the reaction times, we considered the peak day of discussion in terms of volume of tweets, counting from the day the attack ended. We found that the *delay in achieving the peak varies significantly between the events*. For cryptocurrencies that have been attacked multiple times, the reaction time is longer for the first event. *For later events, the users are more alert and hence the peak days are closer to the day the attack ended.*

To study the sentiments of users around the events in the timeline, we have defined the *sentiment profile* for a dataset to capture the percentage distribution of positive, neutral and negative sentiments as categorised by the composite score of VADER (Valence Aware Dictionary for sEntiment Reasoning) lexicon [40]. The lexicon-based approach in place of heavier (machine learning based) techniques was necessary because many of our datasets are quite small. When the cryptocurrencies are not under attack, we found some general variation in the sentiment profile of the datasets. While some cryptocurrencies like Bitcoin, Litecoin Cash and Vertcoin have a high degree of positive and neutral sentiments when not under attack, that can not be generalised for all cryptocurrencies on our timeline. *When under attack, however, almost every cryptocurrency succumbs to negative sentiments.*

For more fine-grained analysis, we have defined the *emotion profile* for a dataset. Each tweet is individually profiled with normalised intensities of five emotions - happy, sad, surprise, fear and anger, as characterised by the Text2-Emotion lexicon [41]. This essentially distributes the tweets into 32 sets characterised by the presence or absence of the five emotions. The emotion profile of a dataset has been defined in terms of the volume and averages of the intensities of the emotions and their combinations. The comparison between the emotion profiles of the three types of datasets for an event as well as across events provides some general trends and deeper insights for some specific cryptocurrencies. In particular, we note that *fear is a perpetually dominant emotion with the highest mean intensity across all cryptocurrencies. However, at the time of the attack, the most consistent observation is the reduction in both volume and intensity of happy emotions, coupled with an intensification of angry emotions.* This observation could be used as a trigger in an attack detection system.

We found other common patterns across cryptocurrencies through intra-currency comparisons. First, we found a common "attack period" of 6 days for all events when there is a high volume of related tweets. When a cryptocurrency is attacked more than once, the peak day in terms of the volume of tweets is quicker for subsequent attacks (Fig. 3). We have presented inter-currency comparisons through pictorial views – Fig. 1 comparing sentiments, and Figs. 2, 12, 13, 14, 15 and 16 comparing the volume and intensities of the five emotions. We have found consistent patterns across not just the largest PoW cryptocurrencies and different periods of time, but for the diverse set of cryptocurrencies included in our analysis, hence *demonstrating generalisability of our conclusions*.

Our lexicon-based analysis of social media is a general technique applicable even on small datasets, and the common patterns that have emerged could be used to deploy lightweight triggering systems for detecting attacks on cryptocurrencies using social media. This would be a *low-cost one-point solution for all cryptocurrency attacks*. Our technique relies on an attack being first reported on social media by experts who are observing the underlying blockchain and enthusiasts with strong interest in the attacked cryptocurrency sharing those posts[11]. As elaborated in Section III-A, the use of social media for disseminating such information is common practice, justifying the feasibility of our approach for an effective solution to the *early detection* and *notification* problems. More generally, for any phenomenon on a social media platform with insufficient data (but high impact like in the case of cryptocurrencies worth trillions of dollars), their volumetric, sentiment and emotion profiling can provide very simple yet informative analysis of key considerations like reaction times of users.

In addition to the cryptocurrency community, our study contributes significantly to the field of Computer-Human Interaction (CHI), particularly in understanding how users perceive and react to security threats in digital environments [8], [42]. Our NLP-based approach is similar to methods used in CHI [8], [42] to understand user behaviour and sentiment in response to various digital interactions [28], [43], in exploring user trust and security perceptions in online platforms [44]. *Our use of volumetric and temporal analyses to profile sentiment variations during attacks and non-attack periods introduces novel methodologies that can be applied to other CHI studies.* It enhances the toolkit available to CHI researchers for analysing user interactions and behaviours in response to specific events.

We have made our cleaned datasets available at [45]. Our code is available at [46].

II. Background

A brief background on blockchains and the 51% attack has been provided in Appendices A and B respectively. Our

[11]Example in Fig. 4

contributions and the results do not depend on these details. A more detailed exposition on blockchains related to this work may be found in the books [11], [47].

A. Sentiment analysis background

Sentiment analysis or opinion mining is a widely used NLP technique with the aim to classify texts into typically three polarity levels, namely negative, neutral, and positive. Such analysis can use machine learning algorithms, which will require a fair amount of labelled training data to be able to classify the rest of the text into sentiment categories. In the absence of labelled training data from the context under investigation, a rule-based lexicon (a dictionary of labelled words and phrases) is often used [48], [49] where words and phrases are already scored against a pre-defined set of associated sentiments by intensity (e.g., VADER or TextBlob). As a result, such lexicon-based techniques are easily explainable. We employed the VADER (Valence Aware Dictionary for sEntiment Reasoning) lexicon [40] which has been used by many studies [50]–[55] with high accuracy on micro-blog texts such as Twitter. Given a text, the VADER sentiment analyser will output a compound score, aggregated from each word's valence in the text. This score varies between -1 indicating that the text is extremely negative and 1 indicating that the text is extremely positive. Along with sentiments, the volumes of social media posts are often considered simultaneously. For example, [56] found that both sentiments and volumes are important factors for stock market price prediction. In contrast, [57] found that for Bitcoin and Ethereum, the tweet volume is a better price predictor than the tweet sentiments, which in their dataset was overwhelmingly positive. We examine both the sentiments and the volumes of tweets as indicators of user perception and reaction.

B. Emotion detection background

Emotion detection is an alternative way to gain insights about people's opinions on topics. It identifies human emotions such as happiness, sadness, surprise, anger and fear from texts. Emotions on social media are known to have significant impact on people's behaviour, especially in making decisions in financial markets [58], [59]. There are some predefined libraries for emotion detection, such as the Text2Emotion lexicon [41] or the NRC Emotion lexicon [60]. As explained above, the outputs of these lexicon-based libraries are easily explainable. The Text2Emotion lexicon outputs five types of emotions whereas the NRC Emotion lexicon outputs eight. However, the former has the advantage of relatively simple setup and easy interpretation and has been used in different contexts with social media data [61]. For example, [62] used it to detect emotions in tweets of the top 150 companies listed in the New York Stock Exchange - specifically, its effects on the price movements of the these companies from the Fortune 1000 list. They found that emotions were "significant predictors" for stock price movements of firms. Similarly, [63] utilised the Text2Emotion lexicon for annotation (labelling) social media data on cryptocurrencies and found that most tweets are associated with happy emotions, followed by fear and surprise. They employed the Text2Emotion library, considering its outputs as ground truth, to later train their neural network models. They reported a higher accuracy of this library than the corresponding machine learning model. The Text2Emotion Python library classifies the text into five emotions – happiness, sadness, surprise, anger and fear – by intensity and their sum is either 0 or 1. In this study we also use the Text2Emotion lexicon library to explore patterns in the emotions of tweets.

III. Methodology

A. Data collection

Our data collection followed a two-step process. We first conducted manual aggregation of 51% attack events on various cryptocurrencies and created a timeline of these events. Thereafter, we collected Twitter data around the identified time periods for each of those events.

1) Creating the Timeline: There is no documented compilation of all 51% attacks on cryptocurrencies, to the best of our knowledge. So we have manually compiled a set of 51% attack events. We call this chronologically ordered set the *timeline of events*. Each event has a *period* (t^s, t^e), where t^s denotes its *start date* and t^e denotes its *end date*. The chronology is determined by the start dates t^s of the events. We have discovered 31 events in total. The *full timeline* is presented in Table I.

There is no established methodology of creating such a timeline for cryptocurrency events. We followed a two-pronged approach for identifying the events and their dates - spotting online artefacts, and corroborating them with social media data. We came to know of many 51% attacks over the years. Additionally, we performed comprehensive secondary research via blockchain news portals and discussion fora followed by general internet search to identify more events. For the events thus identified, we noted the dates when they occurred. This gave us the initial timeline of events. We then conducted experimental searches on Twitter to find tweets related to those events around the noted dates. In particular, we searched with the name and the acronym of a cryptocurrency along with the phrase "51 attack". We manually went through the tweets to check if they indeed corroborated with the attack. Each event in our timeline of Table I has been referenced, while the details of the references are in Table III in the Appendix[12].

Our methodology for constructing the timeline has its limitations. Given the thousands of cryptocurrencies[13] that exist today [2], we may have missed very low profile 51% attacks. Such attacks may not have been reported at all on the internet and hence were not captured in our searches.

While creating the full timeline, we observed the following. The first known 51% attack on a cryptocurrency was on Feathercoin in 2013 [R01, R02]. The first major event concerning a

[12]The references confirming the attack events are indexed as [Rxx] in Table III.

[13]A list is also available at https://coinmarketcap.com/all/views/all/.

possible 51% attack on a cryptocurrency that grabbed popular attention was in 2014 for Bitcoin [R07, R08, R09]. After the huge jump in the prices of major cryptocurrencies like Bitcoin in 2017, there were a wide variety of high profile and low profile 51% attacks. Twitter became a common medium for real-time dissemination of news and opinions on cryptocurrency events (much like for other financial instruments [22]–[26]) since around that time. Influential Twitter profiles such as that of Vitalik Buterin, one of the creators of Ethereum, started tweeting on 51% attacks and other events. Despite the increasing number of reported cases, the *mainstream media has not picked up these incidents*, as is evident from our reference list in Table III, Appendix D. Hence we chose Twitter data for our analysis.

2) Gathering Twitter Data: For our comparative analysis, we selected 17 of the 31 events, labelled with E_i in Table I. We call this shortened list as the *reduced timeline* T. These are events of *cryptocurrencies that were attacked more than once*, so that comparisons between events of the same cryptocurrency is possible. The only exceptions are the events of Feathercoin in 2013 [R01, R02] and Bitcoin in 2014 [R07, R08, R09], included for their significance explained earlier. We have gathered Twitter data for events in the reduced timeline T. We chose an *attack data period* $(t^s - 1, t^e + 6)$ for collecting the data on each $E_i \in T$. This meant that we searched for tweets from *one day before*, until *one week after* the event period. For example, the period of the event E_4 was from 16 May 2018 to 19 May 2018 when Bitcoin Gold was attacked. So we collected tweets for this event for the duration 15 May 2018 (i.e. one day before the start date 16 May 2018 of E_4) until 25 May 2018 (i.e. one week after the end date 19 May 2018 of E_4). We started from $t^s - 1$ to provide a buffer to the start date t^s reported in the references of Table III. Our choice for $t^e + 6$ is based on our observation on the volume of tweets of all events E_i. As an example, consider Fig. 5 in Appendix D for the attack event E_{12} on Bitcoin Gold. It shows that the volume of tweets keeps growing until it reaches a peak and then goes back to almost the pre-attack levels within 6 days from the day the attack ended. This is interestingly true for almost all events, thus providing a rationale for the common period. Also note from Fig. 5 that most tweets are negative during this period. We used the full name of the cryptocurrency (e.g. "Bitcoin Gold" for E_{12}) as the keyword to search for relevant tweets using the Twitter API [64]. For each event $E_i \in T$, we created three datasets as follows.

- *A whole dataset* $\mathcal{W}_i$ for the attack data period $(t_i^s - 1, t_i^e + 6)$,
- *An attack dataset* $\mathcal{A}_i$ that is a subset of the whole dataset $\mathcal{W}_i$, containing tweets *explicitly discussing the attack*, for the attack data period $(t_i^s - 1, t_i^e + 6)$, and
- *A benchmark dataset* $\mathcal{B}_i$ containing all tweets for a *benchmark data period* that is a *month before the period of the event*[14].

[14]The only exceptions are event E_{15} of Ethereum Classic and E_{17} of Bitcoin SV, where the benchmark data were taken a month after the attacks, to avoid coinciding with data periods from earlier attacks.

3) Data Preparation: After creating the *initial raw datasets*, we conducted manual inspection on samples of collected tweets to check their suitability for our studies. We found that the data required some cleaning and in some cases we needed to refine our search to get more appropriate data. For cryptocurrencies like Bitcoin whose names are unambiguous, there was no need to use extra keywords to remove tweets that were off-topic. However, in some cases such as "Verge" or "Expanse", where the name of the cryptocurrency is a general word with other common meanings, additional keywords were used to refine the search and only keep the relevant tweets in the results. These additional keywords were the coin abbreviations (such as XVG and EXP), as well as words related to cryptocurrencies and their mining - "crypto", "coin", "currency", "miner", and "mining". Only those tweets that included at least one of these terms were kept for further analysis. This strategy was used for all three types of datasets - $\mathcal{W}_i$, $\mathcal{A}_i$ and $\mathcal{B}_i$ - for an event $E_i \in T$. The attack datasets $\mathcal{A}_i$ were created as a subset of the whole dataset $\mathcal{W}_i$ only containing tweets that contained reference to 51% attack. So these were tweets containing at least one of the keywords "51", "attack" or "double spend".

We conducted significant manual checks to ensure that our methods do not usually generate false positive or false negative results. We have manually checked 100 random samples of tweets[15] from the Bitcoin Gold dataset $\mathcal{W}_4$ and scored the validity of the chosen libraries. We scored a tweet 1 if the sentiment or emotion was justified, and 0 otherwise. We found that both VADER and Text2Emotion libraries perform with 81% accuracy.

We also cleaned the datasets to leave out non-English tweets and duplicates from the same Twitter handles. We then applied the typical text mining based cleaning steps: removal of (1) url links, (2) special characters, and (3) extra white spaces. We finally converted all alphabets to the lower case.

B. Sentiment profiling

We have used the VADER lexicon's compound scores to classify tweets into positive, neutral and negative sentiments [40]. Our choice of the lexicon-based approach in place of heavier (machine learning based) techniques was necessary because many of our datasets are quite small (like E_6, E_9, E_{16} have only tens or hundreds of tweets). It is also a simple and effective tool whose results are transparent (explainable) and can be easily reproduced. Building upon the classification of individual tweets through their compound scores, we define the following to characterise a dataset.

Definition 3.1: Let $\mathcal{D} = (u_1, \ldots, u_n)$ be a text dataset with n units of text data u_i, each with a compound score of $-1 \leq s_i \leq 1$. Let δ_p denote the threshold above which the score is considered positive and δ_n is the threshold below which the score is considered negative. The *sentiment profile* of $\mathcal{D}$ is defined as

$$SP(\mathcal{D}) = (N(\mathcal{D}), Z(\mathcal{D}), P(\mathcal{D})), \tag{1}$$

[15]In "accuracy_check_sample_BTG_2018.xlsx" in our data repository [45].

where

$$P(\mathcal{D}) = |\{u_i : s_i \geq \delta_p, u_i \in \mathcal{D}\}| \times (100/n), \quad (2)$$

$$Z(\mathcal{D}) = |\{u_i : s_i > \delta_n, s_i < \delta_p, u_i \in \mathcal{D}\}| \times (100/n), \text{ and} \quad (3)$$

$$N(\mathcal{D}) = |\{u_i : s_i \leq \delta_n, u_i \in \mathcal{D}\}| \times (100/n). \quad (4)$$

Here, $N(\mathcal{D}), Z(\mathcal{D})$ and $P(\mathcal{D})$ denote the percentages of u_i's in $\mathcal{D}$ that are negative, neutral and positive respectively. Note that $0 \leq N(\mathcal{D}), Z(\mathcal{D}), P(\mathcal{D}) \leq 100$ and $N(\mathcal{D}) + Z(\mathcal{D}) + P(\mathcal{D}) = 100$. For our analysis, $\delta_n = -0.0005$ and $\delta_p = 0.0005$. Different positive and negative thresholds could be used to tune the system depending on its context.

C. Emotion profiling

The Text2Emotion lexicon scores five emotions – happiness, sadness, surprise, fear and anger – between 0 and 1 such that their sum is 0 when the intensity of each emotion is 0, and 1 otherwise. We are not only interested in the dominant emotion in a dataset, but the variation of all five emotions in the tweets as well (as in [63]). We have profiled the emotions by their *intensity* and *volume*.

For any real number $x \in \mathbb{R}$, *ceiling of* x denoted as $\lceil x \rceil$ is the smallest integer greater than or equal to x. For a data unit $u_i \in \mathcal{D}$, let h_i denote the intensity of the happy emotion, a_i for anger, s_i for surprise, d_i for sadness, and f_i for fear.

Definition 3.2: Let $\mathcal{D} = (u_1, \ldots, u_n)$ be a text dataset with n units of text data. The *emotion intensity* of a data unit u_i is defined as

$$\texttt{ei}(u_i) = (h_i, a_i, s_i, d_i, f_i) \quad (5)$$

such that $0 \leq h_i, a_i, s_i, d_i, f_i \leq 1$ and

$$h_i + a_i + s_i + d_i + f_i = \begin{cases} 0, & \text{if } h_i = a_i = s_i = d_i = f_i = 0; \\ 1, & \text{otherwise.} \end{cases} \quad (6)$$

The emotion intensity of the dataset $\mathcal{D}$ is defined as

$$\texttt{EI}(\mathcal{D}) = (H(\mathcal{D}), A(\mathcal{D}), S(\mathcal{D}), D(\mathcal{D}), F(\mathcal{D})) \quad (7)$$

such that

$$H(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} h_i, A(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} a_i, S(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} s_i, \quad (8)$$

$$D(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} d_i, \text{ and } F(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} f_i. \quad (9)$$

The emotion volume of a data unit u_i is defined as

$$\texttt{ev}(u_i) = (hv_i, av_i, sv_i, dv_i, fv_i) \quad (10)$$

such that

$$hv_i = \lceil h_i \rceil, av_i = \lceil a_i \rceil, sv_i = \lceil s_i \rceil, dv_i = \lceil d_i \rceil, fv_i = \lceil f_i \rceil. \quad (11)$$

The emotion volume of the dataset $\mathcal{D}$ is defined as

$$\texttt{EV}(\mathcal{D}) = (HV(\mathcal{D}), AV(\mathcal{D}), SV(\mathcal{D}), DV(\mathcal{D}), FV(\mathcal{D})) \quad (12)$$

such that

$$HV(\mathcal{D}) = \sum_{i=1}^{n} hv_i, AV(\mathcal{D}) = \sum_{i=1}^{n} av_i, SV(\mathcal{D}) = \sum_{i=1}^{n} sv_i, \quad (13)$$

$$DV(\mathcal{D}) = \sum_{i=1}^{n} dv_i, \text{ and } FV(\mathcal{D}) = \sum_{i=1}^{n} fv_i. \quad (14)$$

The emotion profile of a dataset $\mathcal{D}$ is defined as

$$\texttt{EP}(\mathcal{D}) = (\texttt{EI}(\mathcal{D}), \texttt{EV}(\mathcal{D})). \quad (15)$$

For a data unit (i.e. a tweet) u_i, even if only one of h_i, a_i, s_i, d_i, f_i has a non-zero value, then $h_i + a_i + s_i + d_i + f_i = 1$. If there are two non-zero emotions, each carrying the same intensity, their values would be 0.5 each, and so on. For a dataset $\mathcal{D}$, $\texttt{EI}(\mathcal{D})$ denotes its emotion intensity. Each element in the tuple $\texttt{EI}(\mathcal{D})$ is an average of the intensities of a corresponding emotion in all tweets in the dataset: $H(\mathcal{D})$ is the average for happy, $A(\mathcal{D})$ is for anger, $S(\mathcal{D})$ is for surprise, $D(\mathcal{D})$ is for sadness, and $F(\mathcal{D})$ is for fear. A dataset gets characterised by these mean values in its emotion intensity. The emotion volume $\texttt{EV}(\mathcal{D})$ of the dataset has counts of the number of data units $u_i \in \mathcal{D}$ containing the five emotions: $HV(\mathcal{D})$ is the number for happy, $AV(\mathcal{D})$ is for angry, $SV(\mathcal{D})$ is for surprise, $DV(\mathcal{D})$ is for sadness, and $FV(\mathcal{D})$ is for fear. The emotion profile $\texttt{EP}(\mathcal{D})$ of the dataset is the pair $(\texttt{EI}(\mathcal{D}), \texttt{EV}(\mathcal{D}))$. We compare the events $E_i \in T$ using the emotion profile of their datasets $\mathcal{W}_i, \mathcal{B}_i$ and $\mathcal{A}_i$.

IV. The Results

A. Timeline of 51% attacks

The 31 attacks we discovered are arranged chronologically in Table I along with references to news articles and other relevant weblinks that corroborate their occurrences. These references are listed in Table III in the Appendix. The attacks are on 20 different cryptocurrencies, as listed in Table II. The events for the top 8 cryptocurrencies in Table II have been included in our analysis of the reduced timeline T. The first 6 of them have been attacked more than once, allowing us to compare between attacks on the same cryptocurrency. Ethereum Classic tops the list with 4 attacks. The attacks on Feathercoin and Bitcoin have been included in T because of their special significance as explained in Section III-A1. The column titled "Event" in Table I denotes the reduced timeline T as a set of events $E_1, \ldots, E_{17}$. Hence our analysis is on $\#T = 17$ events and 51 datasets - three datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$, for each event $E_i \in T, 1 \leq i \leq \#T$. The event E_4 is when Bitcoin faced only a threat of a 51% attack, but was not actually attacked. The event E_{12} is for an attack on Bitcoin Gold that was averted. The columns titled $|\mathcal{W}_i|$, $|\mathcal{A}|$ and $|\mathcal{B}_i|$ contain the number of tweets in the whole dataset $\mathcal{W}_i$, the attack dataset $\mathcal{A}_i$ and the benchmark dataset $\mathcal{B}_i$ respectively, for each $E_i \in T$. Comparing the volumes $|\mathcal{W}_i|, |\mathcal{A}_i|$ and $|\mathcal{B}_i|$ we see that in most cases, the number of tweets increase during the attack in comparison to the period when there is no attack. At the time of the attack, most of the tweets are related to the

attack. This reaffirms our understanding that *those interested in a cryptocurrency are certainly vigilant about its security against the* 51% *attack as well*. For cryptocurrencies that have been attacked or threatened multiple times, *the first event had the largest volume, and it kept decreasing in subsequent events*. This is also true for benchmark datasets, *indicating a reduced enthusiasm in the respective cryptocurrencies, after attacks*.

B. Comparing sentiment profiles

The sentiment profile $\texttt{SP}(\mathcal{D})$ of a dataset $\mathcal{D}$ captures the percentages of tweets categorised into positive, neutral and negative sentiments. Fig. 1 presents the sentiment profiles $\texttt{SP}(\mathcal{D})$ of the datasets $\mathcal{D} = \mathcal{W}_i$, $\mathcal{B}_i$ and $\mathcal{A}_i$ respectively, for all $E_i \in T$. The sentiment profiles of the benchmark datasets $\texttt{SP}(\mathcal{B}_i)$ in Fig. 1.C are much more positive with fewer negative tweets in them, when compared with $\texttt{SP}(\mathcal{W}_i)$ in Fig. 1.A or $\texttt{SP}(\mathcal{A}_i)$ in Fig. 1.B. So the 51% attacks are significantly noticeable events on Twitter with *a clear change in the proportions of sentiments from positive/neutral to negative* for most cryptocurrencies. Hence sentiment profiles could be used for *the effective detection of deviation from the "norm"* and lead to the design of a *triggering mechanism for flagging such events*.

We additionally observe a clear difference in the sentiment profiles of the whole datasets $\mathcal{W}_i$ and attack datasets $\mathcal{A}_i$. In most cases the proportion of negative tweets are significantly higher in $\mathcal{A}_i$ than in $\mathcal{W}_i$, while positive tweets are nearly eliminated from $\mathcal{A}_i$. For example, the positive sentiments of the first known attack event E_1 on Feathercoin [8 June 2013] dropped from 32.57% in $\mathcal{W}_1$ to 3.85% in $\mathcal{A}_1$. The neutral tweets generally follow the same pattern of decrease in $\mathcal{A}_i$ from $\mathcal{W}_i$. However, we found two events (both for Litecoin Cash) where the positive sentiments dropped, but the main sentiment was neutral and not negative, contrary to all other events. This could indicate that users are generally very positive about the cryptocurrency and are unfazed by 51% attacks. For the event E_2 of Bitcoin in 2014, the positive sentiments are significantly higher than in any other event. This could be primarily because this was only a threat and not an actual attack. It may also indicate the high level of trust and support in the cryptocurrency community where users are positively passionate about the first and most popular cryptocurrency Bitcoin.

C. Comparing emotion profiles

The emotion profile $\texttt{EP}(\mathcal{D})$ for a dataset $\mathcal{D}$ is characterised by five emotions - happiness, sadness, surprise, fear and anger. It contains the emotion volume $\texttt{EV}(\mathcal{D})$ denoting the numbers of tweets in $\mathcal{D}$ that carried each of the five emotions, and the emotion intensity $\texttt{EI}(\mathcal{D})$ denoting the average intensities of each of the five emotions.

The *emotion volumes* $\texttt{EV}(\mathcal{W}_i), \texttt{EV}(\mathcal{B}_i)$ and $\texttt{EV}(\mathcal{A}_i)$ are represented as heat maps A, B and C in Fig. 2 respectively. The columns in a heat map corresponds to the events $E_i \in T$. A row represents a combination of the five emotions. The last row labelled $(0, 0, 0, 0, 0)$ is for *neutral* tweets in which none of the five emotions were detected. For any other row, its label indicates the emotions with positive counts therein. For example, the top row labelled $(HV, -, -, -, -)$ is for tweets in which only the happy emotion was detected. The row labelled $(-, -, SV, -, FV)$ is for tweets that carry only surprise and fear. A cell in the heat map is for the tweets that carry only the emotions indicated in its column label, for the event E_i denoted in its column label, in the dataset ($\mathcal{W}_i$, $\mathcal{A}_i$ or $\mathcal{B}_i$) denoted below the heat map. The colour of the cell denotes the percentage of tweets as indicated in its legend to the right. Darker colours denote higher percentages. A tweet gets counted in a cell if the emotions in its row label have been detected in it, regardless of the magnitudes of the emotions in the tweet. The column for E_i in Fig. 2.A (respectively Figs. 2.B and 2.C) is essentially a partition of the total count of tweets in the dataset $\mathcal{W}_i$ (respectively $\mathcal{A}_i$ and $\mathcal{B}_i$) into the counts of tweets carrying the specific emotions indicated in the respective row labels. So the counts in the cells of each column add up to 100%.

Our observations from the heat maps are as follows. For majority of the tweets in almost every $E_i \in T$, no emotions have been detected and hence they have been counted in the column labelled $(0, 0, 0, 0, 0)$. Hence the last row is the darkest in each of the heat maps in Fig. 2 (A, B and C). However, when emotions are expressed, they are mostly one at a time in a tweet. Hence the first five rows representing tweets with singleton emotions are the next darkest ones. Among these first five rows, the fifth row from the top for fear is usually darkest, indicating it to be the most dominant singleton emotion. The fifth row from the top is also generally darker in Fig. 2.A than in Fig. 2.C, denoting that *fear grows at the time of a* 51% *attack on a cryptocurrency, compared to when it is not being attacked.* When emotions occur in pairs or when more than two emotions are present in a tweet, fear is also usually one of them. In the attack datasets $\mathcal{A}_i$ (Fig. 2.B), some interesting patterns could be observed that differ from the whole datasets $\mathcal{W}_i$. Generally, the emotions are more intensely present in $\mathcal{A}_i$. Fear is the most dominant emotion in most $\mathcal{W}_i$, while there are some significant shifts toward anger, sadness and neutrality in $\mathcal{A}_i$. In benchmark datasets $\mathcal{B}_i$ (Fig. 2.C), happiness is more significant both as a singleton emotion and in various combinations with the other emotions, while anger is hardly present at all. So, *people are generally happy with cryptocurrencies, until they are attacked.*

There are some deviations in our general observations. For event E_{11} on Bitcoin Gold [23-24 Jan 2020], within the whole dataset $\mathcal{W}_{11}$, the most dominant emotion expressed is anger, followed by no emotions in neutral tweets and then the pair (sadness, fear). It is also striking that Litecoin Cash whose sentiment profile in $\mathcal{A}_i$ previously showed that it mainly has neutral sentiments, still has 75% of neutral tweets with no emotions in the first instance ($\mathcal{A}_6$), whereas during the second attack ($\mathcal{A}_9$), fear is overwhelmingly dominant at 84.48%, refuting the neutrality of the sentiment profile. *This demonstrates the advantage of the granularity in capturing reactions through the emotion profile over the sentiment profile.*

TABLE I

THE FULL TIMELINE OF ALL 51% ATTACK EVENTS WE IDENTIFIED, WITH THEIR ATTACK PERIODS (t^s, t^e), PEAK DAYS AND DATASET SIZES. EVENTS 'E_i' ARE PART OF OUR ANALYSIS; OTHERS ARE LABELED '-'. ** THESE EVENTS WERE THREATS OR ATTACKS THAT WERE AVERTED.

Sl.No.	Event	Cryptocurrency	Attack Period (t^s, t^e)	Peak day	$\lvert\mathcal{W}_i\rvert$	$\lvert\mathcal{A}_i\rvert$	$\lvert\mathcal{B}_i\rvert$
01	E_1	Feathercoin [R01, R02]	(08 Jun'13, 08 Jun'13)	10 Jun	261	26	51
02	-	Powercoin [R03, R04, R05]	(15 Jun'13, 16 Jun'13)				
03	-	Terracoin [R03, R06]	(25 Jul'13, 25 Jul'13)				
04	E_2	**Bitcoin [R07, R08, R09]	(12 Jun'14, 13 Jun'14)	16 Jun	72458	3409	65408
05	-	Shift [R03, R10]	(25 Aug'16, 25 Aug'16)				
06	-	Krypton [R03, R10, R11]	(26 Aug'16, 26 Aug'16)				
07	-	Electroneum [R03, R12]	(04 Apr'18, 04 Apr'18)				
08	E_3	Verge [R13, R14, R15, R16]	(04 Apr'18, 04 Apr'18)	05 Apr	6155	599	7804
09	E_4	Bitcoin Gold [R17, R18]	(16 May'18, 19 May'18)	24 May	2449	1012	1137
10	-	MonaCoin [R03, R19]	(17 May'18, 17 May'18)				
11	E_5	Verge [R20, R21, R22]	(21 May'18, 22 May'18)	23 May	3661	685	4030
12	E_6	Litecoin Cash [R23]	(30 May'18, 30 May'18)	05 Jun	842	28	1338
13	-	ZenCash [R24, R25, R26]	(02 Jun'18, 03 Jun'18)				
14	-	FLO [R06, R27]	(08 Sep'18, 08 Sep'18)				
15	-	Pigeoncoin [R03, R28]	(26 Sep'18, 26 Sep'18)				
16	-	Bitcoin Private [R03, R29]	(19 Oct'18, 19 Oct'18)				
17	-	Karbo [R03, R30]	(10 Nov'18, 10 Nov'18)				
18	-	AurumCoin [R06, R31, R32]	(11 Nov'18, 11 Nov'18)				
19	E_7	Vertcoin [R33, R34]	(12 Oct'18, 02 Dec'18)	04 Dec	1596	382	297
20	E_8	Ethereum Classic [R35, R36]	(05 Jan'19, 08 Jan'19)	08 Jan	6840	4465	1105
21	E_9	Litecoin Cash [R37]	(04 Jul'19, 07 Jul'19)	11 Jul	462	58	542
22	-	Expanse [R38]	(29 Jul'19, 29 Jul'19)				
23	E_{10}	Vertcoin [R39, R40, R41]	(01 Dec'19, 01 Dec'19)	02 Dec	457	378	65
24	E_{11}	Bitcoin Gold [R42, R43, R44]	(23 Jan'20, 24 Jan'20)	27 Jan	941	594	293
25	E_{12}	**Bitcoin Gold [R45, R46]	(02 Jul'20, 10 Jul'20)	11 Jul	712	218	482
26	E_{13}	Ethereum Classic [R47, R48, R49]	(29 Jul'20, 01 Aug'20)	01 Aug	1691	900	246
27	E_{14}	Ethereum Classic [R50, R51]	(05 Aug'20, 05 Aug'20)	06 Aug	1649	967	215
28	E_{15}	Ethereum Classic [R52, R53]	(29 Aug'20, 29 Aug'20)	30 Aug	1121	706	172
29	-	Firo [R54, R55, R56]	(20 Jan'21, 20 Jan'21)				
30	E_{16}	Bitcoin SV [R57]	(24 Jun'21, 09 Jul'21)	11 Jul	2174	96	1625
31	E_{17}	Bitcoin SV [R58, R59]	(03 Aug'21, 03 Aug'21)	04 Aug	1159	524	763

TABLE II

THE NUMBER OF ATTACKS PER CRYPTOCURRENCY AS RECORDED IN TABLE I.

Sl. No.	Cryptocurrency	Number of attacks	Sl. No.	Cryptocurrency	Number of attacks
01	Ethereum Classic	4	09	Bitcoin Private	1
02	Bitcoin Gold	3	10	Electroneum	1
03	Bitcoin SV	2	11	Expanse	1
04	Litecoin Cash	2	12	Firo	1
05	Verge	2	13	Karbo	1
06	Vertcoin	2	14	Krypton	1
07	Bitcoin*	1	15	MonaCoin	1
08	Feathercoin*	1	16	Pigeoncoin	1
			17	Powercoin	1
			18	Shift	1
			19	Terracoin	1
			20	ZenCash	1

* Cryptocurrencies that were attacked only once, and yet have been included in our analysis.

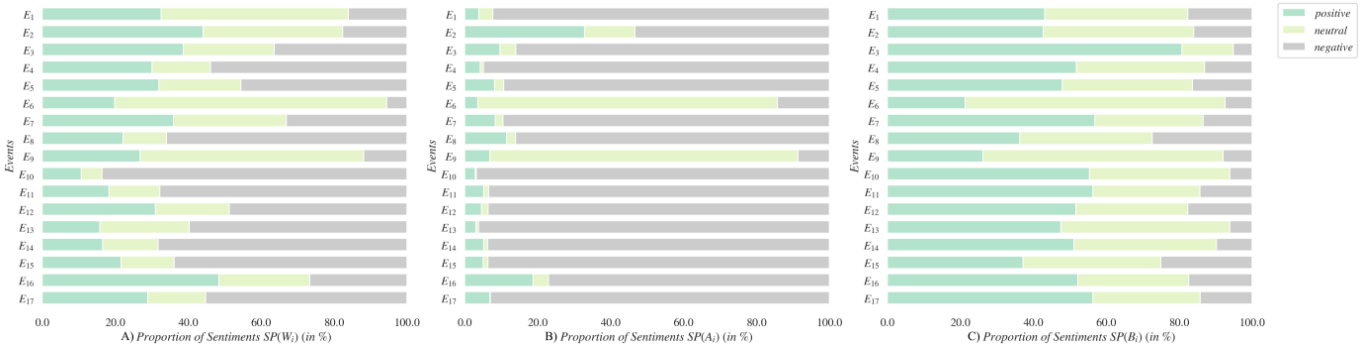


Fig. 1. SP($\mathcal{W}_i$): Sentiment profiles of the A) whole datasets $\mathcal{W}_i$, B) attack datasets $\mathcal{A}_i$ and C) benchmark datasets $\mathcal{B}_i$. Higher resolution images of Fig. 1(A, B and C) in Appendix D as Figs. 6, 7 and 8 respectively.

We next consider the *emotion intensities* for the datasets $\mathcal{W}_i, \mathcal{B}_i$ and $\mathcal{A}_i$. Recollect that the emotion intensity of the

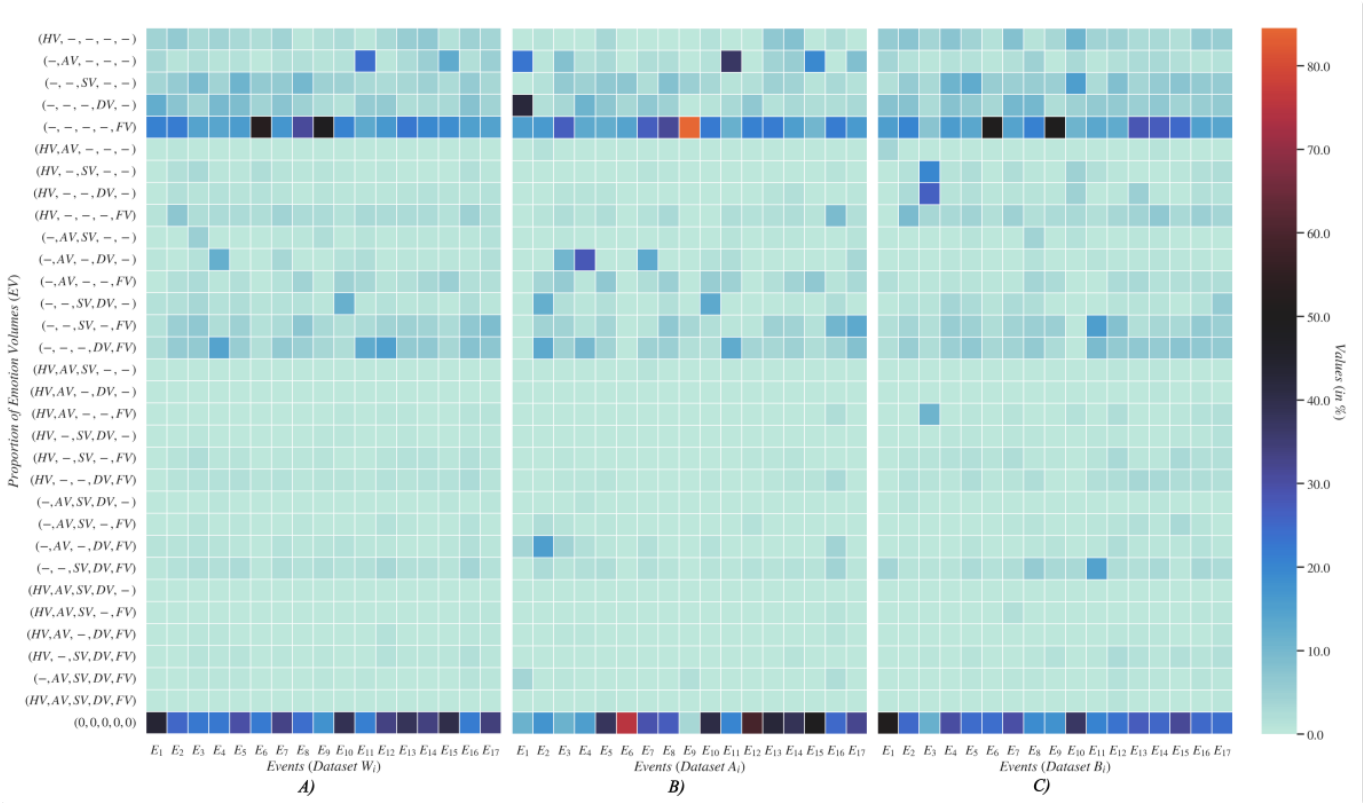


Fig. 2. Heat map of percentages of tweets carrying different emotions and their combinations in A for $\mathcal{W}_i$, B for $\mathcal{A}_i$ and C for $\mathcal{B}_i$ datasets respectively. Higher resolution images of Fig. 2(A, B and C) in Appendix D as Figs. 9, 10 and 11 respectively.

dataset $\mathcal{D}$ is given by

$$\texttt{EI}(\mathcal{D}) = (H(\mathcal{D}), A(\mathcal{D}), S(\mathcal{D}), D(\mathcal{D}), F(\mathcal{D})).$$

These mean intensities have been plotted for comparison in Figs. 12, 13, 14, 15 and 16 respectively. As expected, anger, sadness and fear are more intense in the attack datasets $\mathcal{A}_i$; happiness is generally more intense in the benchmark datasets $\mathcal{B}_i$. The plot for happiness (Fig. 12) shows a good level of consistency. In almost all attack datasets $\mathcal{A}_i$ their means are lower than those in $\mathcal{W}_i$ and $\mathcal{B}_i$ with only the exceptions of the second and third attack on Ethereum Classic (events E_{13} and E_{14}), while anger (Fig. 13) is more intense. The intensity of surprise (Fig. 14) generally averages below 0.2 across all datasets. There is of course a noticeable high peak at E_2 (the Bitcoin event) in the whole dataset. This event shook up the cryptocurrency community and hence the element of surprise. However, not all emotions across the datasets show a consistent pattern. Sadness (Fig. 15) and fear (Fig. 16) have generally varied across datasets. Approximately half of the time in $\mathcal{A}_i$, their intensities are either under or above compared to $\mathcal{W}_i$ and $\mathcal{B}_i$. These two emotions therefore do not give a consistent picture across cryptocurrencies.

Fear has the highest mean intensity across all datasets, compared to the other four emotions. It fluctuates roughly between 0.2 and 0.9, while the others have their means around 0.1. Studying fear closely for the various cryptocurrencies, we see that Bitcoin, Vertcoin and Bitcoin SV have their mean intensities for $\mathcal{A}_i$ above $\mathcal{W}_i$ and $\mathcal{B}_i$ – so users are more fearful at the time of the attack. On the other hand, Feathercoin, Bitcoin Gold and Ethereum Classic have their mean intensities for $\mathcal{A}_i$ slightly under $\mathcal{W}_i$ and $\mathcal{B}_i$ – so users are less fearful at the time of the attack. The third cluster is of Verge and Litecoin Cash that have fluctuating intensities in subsequent attacks. It is also interesting to note that in case of the second attack on Litecoin Cash (E_9) there is a high peak fear intensity of around 0.9. Note that the sentiment profile had only shown that a vast majority of the tweets are neutral. *This further demonstrates that the more granular emotion profile provides better insights than the sentiment profile.* Moreover, these currency-specific observations could indicate *behavioural differences between the various cryptocurrency communities and their supporters*.

D. Analysis of peak days

We further analyse the events $E_i \in T$ by identifying the peak days in terms of the number of tweets, following its attack period (t^s, t^e). We count the peak day starting from the last day of the attack t^e as day 0. Table I provides the peak days for each $E_i \in T$ in the respective $\mathcal{A}_i$. Fig. 3 denotes the distances of the peak days from day 0. Here, the timeline is presented as quarter of year blocks and the cryptocurrencies are colour coded and placed in the timeline. We have observed that the sentiments on the peak days are in

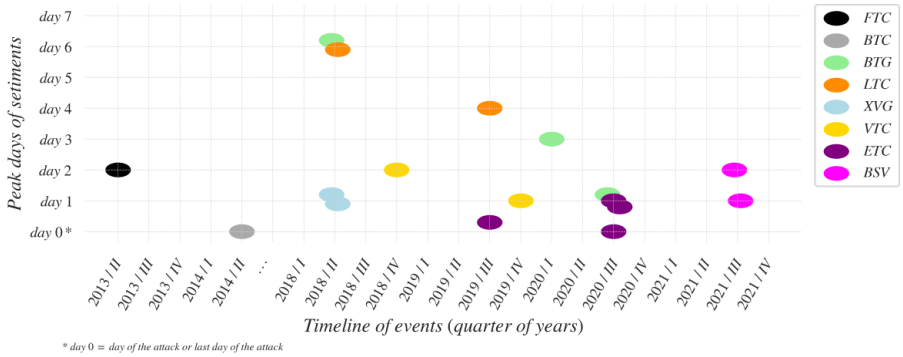


Fig. 3. Peak day delays

most cases overwhelmingly negative. Interestingly, *there is a reaction delay in attaining the peak* from day 0. In most cases, the peak is delayed by a day or two. Sometimes a quick same day reaction (as identified through the time in the tweet) has the same delay as one from the next day because of variations in time-zones. There could be other factors like the time of the day when the attack is first identified and reported. If the information came out in the late hours of the day anywhere on earth, it would fall into the next day category. This anomaly could be observed for example in the case of attacks on more popular cryptocurrencies like Ethereum Classic which have enough data for the anomaly to be identified and distinguished. *In majority of the cases, the reaction delays were initially higher, but significantly decreased in subsequent attacks.* For example, in the first attack on Bitcoin Gold (BTG), the peak was on the sixth day after the attack ended on 19 May 2018. However, for the second attack, the peak was on the third day after 24 January 2020; for the third attempted attack, the peak was on just the next day after 10 July 2020. Another example is Litecoin Cash (LTC) where the peak was attained after the sixth day for the first attack in 2018, while in 2019 it decreased to the fourth day. This reaction delay could be understood by the *Stroop effect* that is very well known in behavioural studies [65].

V. Conclusion

The security of a blockchain (underlying a cryptocurrency) against the 51% attack is fundamental to its integrity. In this work, we have created a first-of-its-kind timeline of 31 such attack events on various proof-of-work cryptocurrencies, since the introduction of Bitcoin in 2009, until August 2021. It reveals the vulnerability of proof-of-work cryptocurrencies in the face of frequent 51% attacks, thus informing against the common perception that such attacks are rather rare. Larger cryptocurrencies are particularly targeted multiple times.

We have constructed tweet datasets around 17 of these attack events and characterised them using lexicon-based tools – VADER lexicon for sentiment analysis and Text2Emotion lexicon for more fine-grained emotion detection. Our analysis shows that for most cryptocurrencies, social media activity significantly increases at the time of the attacks compared to normal periods; They are also filled with overwhelmingly negative sentiments during attacks. More granular analysis shows that fear is generally the predominant emotion at all times, for all cryptocurrencies. However, at the time of attacks, the volume and intensity of happiness decreases, while anger increases in intensity consistently.

The consistent patterns that have emerged out of our analysis show that the volume and perceptions of tweets could be used to deploy a triggering system to alert users of an ongoing attack, remedying the complete lack of such notifications [34].

The detection of attacks and the profiling of user perceptions around them are important for risk analysis of cryptocurrencies. Services like the CertiK Skynet [15] use undisclosed and perhaps heuristic techniques to measure user perceptions and report the "community trust" scores for cryptocurrencies. Our lexicon-based approach for profiling social media data could be used to develop an aggregated score to work as an open-sourced counterpart to their score. We hope to have set the stage for such open techniques for risk analysis that are easily understood, subject to public scrutiny and critique, and hence more explainable and trustworthy.

Finally, many governments around the world have called for regulations on cryptocurrencies[16]. Risk analysis of cryptocurrencies using open techniques have the potential to enable and/or complement such regulatory frameworks by informing the users about the risks in real-time while they are invested in cryptocurrencies. This work is a concrete step in that direction.

There are several directions for immediate research on top of this first study. The first thing to do would be to use our techniques to detect attacks after August 2021. One may also expand the data sources beyond X to Reddit, Discord, Facebook, Instagram, public Telegram/WhatsApp channels, etc. Distinguishing between social media posts by humans and those by bots is essential to capture the truly dominant human sentiments and those amplified by bots. When enough data is available, variations of user perceptions with geography, communities, influencers, etc. can provide deeper insights about the attack events. Filtering out mis-/dis-information is important to ensure robustness of the attack detection process.

[16]UK's promise: https://www.bbc.com/news/technology-64468617

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] Statista, "Cryptocurrencies worldwide from 2013 to january 2025," 2025. https://www.statista.com/statistics/863917/number-crypto-coins-tokens/, last accessed on 07 March 2025.

[3] M. Perry and J. Ferreira, "Moneywork: Practices of use and social interaction around digital and analog money," *ACM Trans. Comput.-Hum. Interact.*, vol. 24, Jan. 2018.

[4] L. Glomann, M. Schmid, and N. Kitajewa, "Improving the blockchain user experience - an approach to address blockchain mass adoption issues from a human-centred perspective," in *Advances in Artificial Intelligence, Software and Systems Engineering* (T. Ahram, ed.), (Cham), pp. 608–616, Springer International Publishing, 2020.

[5] M. Fröhlich, F. Gutjahr, and F. Alt, "Don't lose your coin! investigating security practices of cryptocurrency users," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, (New York, NY, USA), p. 1751–1763, Association for Computing Machinery, 2020.

[6] M. Froehlich, M. R. Wagenhaus, A. Schmidt, and F. Alt, "Don't stop me now! exploring challenges of first-time cryptocurrency users," in *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, DIS '21, (New York, NY, USA), p. 138–148, Association for Computing Machinery, 2021.

[7] M. Froehlich, F. Waltenberger, L. Trotter, F. Alt, and A. Schmidt, "Blockchain and cryptocurrency in human computer interaction: A systematic literature review and research agenda," in *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, (New York, NY, USA), p. 155–177, Association for Computing Machinery, 2022.

[8] D. Murray-Rust, C. Elsden, B. Nissen, E. Tallyn, L. Pschetz, and C. Speed, "Blockchain and beyond: Understanding blockchains through prototypes and public engagement," *ACM Transactions on Computer-Human Interaction*, vol. 29, no. 5, pp. 1–73, 2023.

[9] US Government, The White House, "Fact Sheet: President Donald J. Trump Establishes the Strategic Bitcoin Reserve and U.S. Digital Asset Stockpile," Mar. 2025. https://www.whitehouse.gov/fact-sheets/2025/03/fact-sheet-president-donald-j-trump-establishes-the-strategic-bitcoin-reserve-and-u-s-digital-asset-stockpile/.

[10] D. Birnbaum, ""Trump Establishes Strategic Bitcoin Reserve: What It Means For Markets"," Mar. 2025. https://www.forbes.com/sites/davidbirnbaum/2025/03/06/bitcoin-reserve-created-by-executive-order-america-great-again/.

[11] A. Narayanan, J. Bonneau, E. W. Felten, A. Miller, and S. Goldfeder, *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press, 2016.

[12] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," in *Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014*, pp. 436–454, 2014.

[13] J. Garay, A. Kiayias, and N. Leonardos, "The bitcoin backbone protocol: Analysis and applications," in *Advances in Cryptology - EUROCRYPT 2015* (E. Oswald and M. Fischlin, eds.), (Berlin, Heidelberg), pp. 281–310, Springer Berlin Heidelberg, 2015.

[14] S. Bhattacherjee and P. Sarkar, "Voting games to model protocol stability and security of proof-of-work cryptocurrencies," in *Decision and Game Theory for Security (GameSec'22)* (F. Fang, H. Xu, and Y. Hayel, eds.), (Cham), pp. 297–318, Springer International Publishing, 2023.

[15] CertiK, "Skynet: Security leaderboard - cryptocurrency," 2024. https://skynet.certik.com/leaderboards/security, last accessed on 07 March 2025.

[16] D. Freeman, T. McWilliams, S. Bhattacharyya, C. Hall, and P. Peillard, "Enhancing trust in the cryptocurrency marketplace: A reputation scoring approach," *SMU Data Science Review*, vol. 1, no. 3, p. 5, 2018.

[17] V. Marella, B. Upreti, J. Merikivi, and V. K. Tuunainen, "Understanding the creation of trust in cryptocurrencies: The case of Bitcoin," *Electronic Markets*, vol. 30, no. 2, pp. 259–271, 2020.

[18] P. M. Krafft, N. Della Penna, and A. S. Pentland, "An experimental study of cryptocurrency market dynamics," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), p. 1–13, Association for Computing Machinery, 2018.

[19] L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on Twitter," *Journal of Research in Personality*, vol. 46, no. 6, pp. 710–718, 2012.

[20] J. Dyer and B. Kolic, "Public risk perception and emotion on Twitter during the Covid-19 pandemic," *Applied Network Science*, vol. 5, no. 1, p. 99, 2020.

[21] M. Park, D. McDonald, and M. Cha, "Perception differences between the depressed and non-depressed users in Twitter," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, pp. 476–485, Aug. 2021.

[22] S. Y. Yang, S. Y. K. Mo, and A. Liu, "Twitter financial community sentiment and its predictive relationship to stock market movement," *Quantitative Finance*, vol. 15, no. 10, pp. 1637–1656, 2015.

[23] D. O. Afanasyev, E. Fedorova, and S. Ledyaeva, "Strength of words: Donald Trump's tweets, sanctions and Russia's ruble," *Journal of Economic Behavior & Organization*, vol. 184, pp. 253–277, 2021.

[24] A. Groß-Klußmann, S. König, and M. Ebner, "Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market," *Expert Systems with Applications*, vol. 136, pp. 171–186, 2019.

[25] S. Behrendt and A. Schmidt, "The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility," *Journal of Banking & Finance*, vol. 96, pp. 355–367, 2018.

[26] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *2019 IEEE fifth international conference on big data computing service and applications (BigDataService)*, pp. 205–208, IEEE, 2019.

[27] S. García-Méndez, F. de Arriba-Pérez, A. Barros-Vila, and F. J. González-Castaño, "Targeted aspect-based emotion analysis to detect opportunities and precaution in financial Twitter messages," *Expert Systems with Applications*, vol. 218, p. 119611, 2023.

[28] H. Paakki, H. Vepsäläinen, A. Salovaara, and B. Zafar, "Detecting covert disruptive behavior in online interaction by analyzing conversational features and norm violations," *ACM Transactions on Computer-Human Interaction*, vol. 31, no. 2, pp. 1–43, 2024.

[29] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, 2021.

[30] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[31] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Predictive sentiment analysis of tweets: A stock market application," in *HCI-KDD 2013, SouthCHI 2013, Maribor, Slovenia, July 1-3, 2013*, pp. 77–88, Springer, 2013.

[32] H. Chen, M. Pendleton, L. Njilla, and S. Xu, "A survey on ethereum systems security: Vulnerabilities, attacks, and defenses," *ACM Comput. Surv.*, vol. 53, 06 2020.

[33] L. Zhou, K. Qin, C. F. Torres, D. V. Le, and A. Gervais, "High-frequency trading on decentralized on-chain exchanges," in *2021 IEEE Symposium on Security and Privacy (SP)*, pp. 428–445, 2021.

[34] Y. Wang, P. Zuest, Y. Yao, Z. Lu, and R. Wattenhofer, "Impact and user perception of sandwich attacks in the defi ecosystem," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[35] L. Zhou, X. Xiong, J. Ernstberger, S. Chaliasos, Z. Wang, Y. Wang, K. Qin, R. Wattenhofer, D. Song, and A. Gervais, "SoK: Decentralized finance (DeFi) attacks," in *2023 IEEE Symposium on Security and Privacy (SP)*, pp. 2444–2461, 2023.

[36] S. Houy, P. Schmid, and A. Bartel, "Security aspects of cryptocurrency wallets—a systematic literature review," *ACM Comput. Surv.*, vol. 56, 08 2023.

[37] S. Shanaev, A. Shuraeva, M. Vasenin, and M. Kuznetsov, "Cryptocurrency value and 51% attacks: Evidence from event studies," *The Journal of Alternative Investments*, vol. 22, no. 3, pp. 65–77, 2019.

[38] R. M. Morrow, Allison, "Trump creates a Strategic Bitcoin Reserve one day ahead of White House crypto summit | CNN Business," Mar. 2025.

[39] C. Ye, G. Li, H. Cai, Y. Gu, and A. Fukuda, "Analysis of security in blockchain: Case study in 51%-attack detecting," in *2018 5th International Conference on Dependable Systems and Their Applications (DSA)*, pp. 15–24, 2018.

[40] C. Hutto and E. Gilbert, "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, pp. 216–225, 2014.

[41] T. Team, "Text2emotion 0.0.5 python library," 2020. https://pypi.org/project/text2emotion/, last accessed on 07 March 2025.

[42] H. Cho, P. Li, and Z. H. Goh, "Privacy risks, emotions, and social media: A coping model of online privacy," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 27, no. 6, pp. 1–28, 2020.
[43] M. C. Caschera, F. Ferri, and P. Grifoni, "Sentiment analysis from textual to multimodal features in digital environments," in *proceedings of the 8th international conference on Management of Digital EcoSystems*, pp. 137–144, 2016.
[44] J. J. Si, T. Sharma, and K. Y. Wang, "Understanding user-perceived security risks and mitigation strategies in the web3 ecosystem," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–22, 2024.
[45] Z. Baruwa and Z. Zhu, "Our cleaned datasets for this work," 2025. https://github.com/zb15/SM_perceptions_51-Attacks/tree/main/Datasets.
[46] Z. Baruwa, "Our code for running the experiments and the results for this work," 2025. https://github.com/zb15/SM_perceptions_51-Attacks.
[47] S. Chakravarty and P. Sarkar, *An introduction to algorithmic finance, algorithmic trading and blockchain*. Emerald Group Publishing, 2020.
[48] T. Inuduka, A. Yokose, and S. Managi, "Noise trader impact: Bitcoin market evidence from Telegram and X," *Social Network Analysis and Mining*, vol. 14, pp. 1–17, Dec. 2024.
[49] S. Kumari, A. Sharma, A. Chhabra, A. Gupta, S. Singh, and R. Verma, "Analysing public sentiment towards robotic surgery: an X (formerly Twitter) based study," *Social Network Analysis and Mining*, vol. 14, pp. 1–18, Dec. 2024.
[50] V. Bonta and N. K. N. Janardhan, "A comprehensive study on lexicon based approaches for sentiment analysis," *Asian Journal of Computer Science and Technology*, vol. 8, no. S2, pp. 1–6, 2019.
[51] A. Borg and M. Boldt, "Using VADER sentiment and SVM for predicting customer response sentiment," *Expert Systems with Applications*, vol. 162, p. 113746, 2020.
[52] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with VADER lexicon polarity detection and k-nearest neighbor algorithm," *Journal of Physics: Conference Series*, vol. 1567, no. 3, p. 032024, 2020.
[53] T. Pano and R. Kashef, "A complete VADER-based sentiment analysis of Bitcoin (BTC) tweets during the era of COVID-19," *Big Data and Cognitive Computing*, vol. 4, no. 4, 2020.
[54] A. Tumasjan, R. Braun, and B. Stolz, "Twitter sentiment as a weak signal in venture capital financing," *Journal of Business Venturing*, vol. 36, no. 2, p. 106062, 2021.
[55] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, p. 102212, 2020.
[56] N. Oliveira, P. Cortez, and N. Areal, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," *Expert Systems with Applications*, vol. 73, pp. 125–144, 2017.
[57] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Science Review*, vol. 1, no. 3, p. 1, 2018.
[58] Y. Ge, J. Qiu, Z. Liu, W. Gu, and L. Xu, "Beyond negative and positive: Exploring the effects of emotions in social media during the stock market crash," *Information Processing & Management*, vol. 57, no. 4, p. 102218, 2020.
[59] J. Griffith, M. Najand, and J. Shen, "Emotions in the stock market," *Journal of Behavioral Finance*, vol. 21, no. 1, pp. 42–56, 2020.
[60] M. M. Bailey, "NRCLex 4.0," 2022. https://pypi.org/project/NRCLex/, last accessed on 07 March 2025.
[61] S. Bhooshan, R. Praveen Pai, and R. Nandakumar, "A sentiment analysis of a boycott movement on Twitter," in *Computer Networks and Inventive Communication Technologies: Proceedings of Fourth ICCNCT 2021*, pp. 313–324, Springer, 2022.
[62] S. Dhar and I. Bose, "Emotions in Twitter communication and stock prices of firms: The impact of Covid-19 pandemic," *Decision*, vol. 47, pp. 385–399, 2020.
[63] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022.
[64] X. (formerly Twitter), "X API v2," 2023. https://developer.x.com/en/docs/x-api, last accessed on 07 March 2025.
[65] C. M. MacLeod, "Half a century of research on the stroop effect: an integrative review.," *Psychological bulletin*, vol. 109, no. 2, p. 163, 1991.
[66] T. Hornyak, "PCWorld article: One group controls 51 percent of Bitcoin mining, threatening security sanctity," 2014. https://www.pcworld.com/article/439835/bitcoin-price-dips-as-backers-fear-mining-monopoly.html, last accessed on 07 March 2025.

APPENDIX

A. Blockchain and cryptocurrency background

A unit of the cryptocurrency is called a *coin*. The blockchain data structure underlying these cryptocurrencies is a hashed chain of blocks that works as a ledger of transactions for the creation and transfer of its coins. To use these coins as a currency, it is necessary that one should not be able to spend the same coin more than once. The ledger is public and is expected to provide an *immutable history* of the transactions, thus preventing multiple spendings of the same coin.

Public blockchains are decentralised and distributed systems. As a data structure, a blockchain is a linear collection of blocks to which new ones can be added at one end. We denote a block as B_i indexed by its position i in the chain

$$\mathtt{B} = (B_0, B_1, \ldots, B_n)$$

of blocks. The first block B_0 is called the *genesis block* and the *block height* of B_i is i.

A *proof-of-work* blockchain involves a cryptographically secure hash function $H : \{0,1\}^* \to \{0,1\}^t$ in the process of finding a new block. A block B_i contains the hash $H(B_{i-1})$ of the previous block B_{i-1} in the chain and a *nonce* η_i such that its own hash $H(B_i)$ has a pre-determined prefix (typically a certain number of 0's) in its t-bit output value. The prefix is determined by the time that was required to generate (some of) the previous blocks in the chain. In order to mine a valid block, a *miner* searches for the appropriate nonce η_i. Finding the nonce is the computational puzzle that gets more and more difficult as the length of the required hash prefix increases. When an appropriate nonce has been found, it is considered as a probabilistic evidence that the miner has done the estimated amount of computational work, and hence the name proof-of-work.

B. Forks and 51% Attacks

Miners usually maintain their own local copy of the entire blockchain data structure. A secure blockchain system promises to provide an immutable history. This means that once a block B_i has been added to the blockchain, the probability that it may be removed from it decreases exponentially as more blocks $B_j, j > i$ are added thereafter.

The possibility that a block that was once added to the chain may get removed from it is due to a *fork* in the chain. A fork occurs when there are two or more chains with common histories. For example,

$$\mathtt{B} = (B_0, B_1, \ldots, B_{i-1}, B_i, B_{i+1}, \ldots, B_n)$$

and

$$\mathtt{B}' = \left(B_0, B_1, \ldots, B_{i-1}, B'_i, B'_{i+1}, \ldots, B'_m\right)$$

are two chains forked from the i^{th} block such that they have the common history $B_0, B_1, \ldots, B_{i-1}$, but then the first chain

has the blocks $B_i, B_{i+1}, \ldots, B_n$ while the second chain has the blocks $B'_i, B'_{i+1}, \ldots, B'_m$.

Forks may occur for various reasons in a blockchain. The creation of new blocks is a randomised event. Being a decentralised system, there is no mechanism of coordination between competing miners in their race to create distinct new blocks. As a result, it is often the case that two or more miners create their own new blocks B_{i+1} and B'_{i+1} on top of the same block B_i at almost the same time. However, the blockchain protocol dictates that *the most difficult (or the longest) chain be considered as valid* at any point of time. This means that only one of these two blocks B_{i+1} and B'_{i+1} makes it to the chain that would be considered valid in the long run. The other block gets *orphaned*. It is possible that the blockchain network gets divided into two parts such that one set of miners attempts to create blocks on top of B_{i+1} while the others attempts to create blocks on top of B'_{i+1}. This results in a more sustained fork. Assuming that all miners choose to follow only one of the two chains, they abandon mining on top of one of the chains and continue mining on the other one. Hence all blocks after the fork in the abandoned chain get orphaned.

A sustained fork may be created with malicious intent. This happens when a set of miners divert from the protocol of building upon the most difficult chain, and instead choose to mine blocks as a fork of the chain followed by the rest of the miners in the network. To launch such an attack successfully on a connected network, the attacker set of miners should possess at least 50% of the resources for mining new blocks. Such an attack is called the 51% *attack*.

A sustained fork may occur when miners disagree on the proposed protocol changes. A set of miners start following a new protocol while the old protocol is followed by the rest of the miners. The blocks generated through one protocol are considered invalid by the other.

A cryptocurrency system uses a blockchain as a ledger of transactions. A unit of the cryptocurrency is called a *coin*. Each block records transactions that either (1) create a certain number of new coins, or (2) transfer those coins from one owner to another. Ownership of coins is ascertained by a cryptographic digital signature system. To transfer the ownership of a coin, a sender uses their secret key to generate a digital signature on a transcript containing the recipient's public key. This signature can be verified using the public key of the sender. An adversary may *hack* into a cryptocurrency owner's wallet (or the exchanges they use) to steal the secret keys. Once they get hold of the secret key or they are able to forge the owner's signature by some other means, they can steal coins from the owner by generating new transactions that subsequently enter the chain. However, this is not an attack on the underlying blockchain. In this work, we study only 51% attacks on cryptocurrency blockchains and not the hacks or other such attacks on cryptocurrencies.

C. Relationship Between Security and Market Value of Cryptocurrencies

There is a clear *interdependence* between the security of a cryptocurrency and its market value. As the value of a cryptocurrency increases, more people "invest" to profit from it, and it becomes more difficult to launch a 51% attack on it. This is easy to see in the case of PoW cryptocurrencies. For example, in Bitcoin, as its market value grows, the miners generally want to increase their computational powers so that their chances of mining new blocks successfully increases and consequently they earn more. As the total computational power invested in the system increases, it becomes more difficult to mine new blocks. At the same time, attacking the system also becomes more difficult. On the other hand, as the computational power of the network decreases – perhaps due to devaluation of the cryptocurrency – the miners move to mine other more profitable cryptocurrencies where they can use the same computational power. Thus, the difficulty of mining new blocks reduces and hence attacking the system becomes easier [R34].

Importantly, the converse is also true. That is, if a cryptocurrency system is threatened by or actually suffers a 51% attack, its value decreases [R59]. In 2014, the mere possibility of a 51% attack on Bitcoin [R07, R08, R09] resulted in an immediate fall in its price [66][17]. In [37], the authors studied 14 such attacks and found that in each case the price of the attacked cryptocurrency immediately decreased by 12% to 15%. We note here that [37] is the only other resource on the internet before our work where a substantial number of 51% attacks have been reported to the best of our knowledge. They also noted that the price does not recover to the pre-attack level for a week after the event. Even though the cryptocurrency market is significantly speculative with various factors affecting the prices, the above result shows the indisputable severity of the 51% attacks in terms of their effect on the prices.

D. Additional Figures and References

This section contains some additional figures. It also contains Table III with online links to reports on attacks, used as references in Table I.

[17]The mining pool `GHash.io` attained more than 51% computational power of the Bitcoin network through its participating miners. Several miners withdrew from the pool to reduce its computational capability and alleviate the possibility of an attack.

TABLE III
REFERENCES FOR THE 51% ATTACKS LISTED IN TABLE I

#	Published by	Article title	Date	URL Link
R01	Feathercoin Forum	FTC 51% attack – Case study	2013	link
R02	TradeBlock Blog	The 51% Attack – What Bitcoin Can Learn From Alt-coin Experiments	2013	link
R03	The Journal of Alternative Investments	Cryptocurrency Value and 51% Attacks: Evidence from Event Studies	2019	link
R04	Bitcointalk	Powercoin 51% Attack	2013	link
R05	Bitcointalk	Confirmation: Powercoin was 51% attacked	2013	link
R06	Komodo Platform Blog	51% Attack Security: Delayed Proof of Work (dPoW)	2018	link
R07	arsTechnica	Bitcoin pool GHash.io commits to 40% hashrate limit after its 51% breach	2014	link
R08	Coindesk	The Bitcoin mining arms race: GHash.io and the 51% issue	2014	link
R09	The Guardian	Bitcoin currency could have been destroyed by '51%' attack	2014	link
R10	Finance Magnates	What is a 51% Attack?	2018	link
R11	DNI	Krypton (KR) suffers a 51% attack in August 2016	2016	link
R12	Finance Magnates	51 Percent Attacks Appear to Have Hit Verge and Electroneum	2018	link
R13	X (formally Twitter) user	Tweet by the handle @vergecurrency.crypto	2018	link
R14	Finder	Verge (XVG) cryptocurrency destroyed in historic 51% attack	2018	link
R15	Finder	Verge cryptocurrency continues dispensing free money despite "fix"	2018	link
R16	Bitcoin News	Verge Is Forced to Fork After Suffering a 51% Attack	2018	link
R17	Bitcoingold Forum	Double Spend Attacks on Exchanges	2018	link
R18	Quartz	Every cryptocurrency's nightmare scenario is happening to Bitcoin Gold	2018	link
R19	NewsBTC	Monacoin Network Still Suffering From Selfless Mining Attack	2018	link
R20	X (formally Twitter) user	Tweet by the handle @$vergecurrency.crypto$	2018	link
R21	The Abacus.io Blog	The Verge Hack, Explained	2018	link
R22	Bitcoin News	Verge Struck by Second PoW Attack in as Many Months	2018	link
R23	CryptoCoinSpy	LiteCoin Cash (LCC) latest victim of a 51% Attack	2018	link
R24	HoriZen.io Blog	ZenCash Statement On Double Spend Attack	2018	link
R25	Crowdfund Insider	Privacy Crypto ZenCash Hacked in 51% Attack	2018	link
R26	Bitcoin News	Bitcoin in brief Monday: Zencash targeted in 51% Attack, Ticketfly hijacked for ransom	2018	link
R27	Medium user	FLO Team Response to 51% attack	2018	link
R28	NewsBTC	Pigeoncoin (PGN) Hacked Due to Bitcoin Protocol Bug, Copycat Coins in Danger?	2018	link
R29	Coindesk	This College Freshman Is Out to 51% Attack Your Cryptocurrency		link
R30	Medium user	Prevent transaction cancellation in 51%-attack	2018	link
R31	Yahoo Finance	Claim: Crypto Exchange Lost $500,000 Due to AurumCoin 51% Attack	2018	link
R32	X (formally Twitter) user	Tweet by Aurum Coin	2018	link
R33	Bitcoin Exchange Guide News	Vertcoin Experiences 51% Blockchain Network Attack as Coinbase Comments About it	2018	link
R34	Coindesk	Vertcoin's Struggle Is Real: Why the Latest Crypto 51% Attack Matters	2018	link
R35	SlowMist	The analysis of ETC 51% attack	2019	link
R36	CoinTelegraph	Ethereum Classic 51% Attack – The Reality of Proof-of-Work	2019	link
R37	Cryptocurrency News	The next 51%: Litecoin Cash (LCC) under attack		link
R38	GitHub Gist user	Expanse (EXP) was 51% attacked	2019	link
R39	GitHub Gist user	Vertcoin (VTC) was 51% attacked	2019	link
R40	Coindesk	The Vertcoin Cryptocurrency Just Got 51% Attacked – Again	2019	link
R41	Bitcoin News	Vertcoin Network Sabotaged by Another 51% Attack	2019	link
R42	gGitHub Gist user	Bitcoin Gold (BTG) was 51% attacked	2020	link
R43	Cointelegraph	Bitcoin Gold Blockchain Hit by 51% Attack Leading to $\$70K$ Double Spend	2020	link
R44	Bitcoin News	Bitcoin Gold 51% Attacked – Network Loses $\$70,000$ in Double Spends	2020	link
R45	Bitcoin Gold	Emergency update 0.17.2	2020	link
R46	Coindesk	Attempted 51% Attack on Bitcoin Gold Was Thwarted, Developers Say	2020	link
R47	HackMD	ETC Chain Split Diagnosis	2020	link
R48	Bitquery	Ethereum Classic 51% Chain Attack July 31, 2020	2020	link
R49	Bitquery	Attacker Stole 807K ETC in Ethereum Classic 51% Attack	2020	link
R50	Coindesk	Ethereum Classic Suffers Second 51% Attack in a Week	2020	link
R51	The Block	Ethereum Classic suffers another 51% attack in five days	2020	link
R52	Coindesk	Ethereum Classic Hit by Third 51% Attack in a Month	2020	link
R53	Bitcoin News	Ethereum Classic Suffers 51% Attack Again: Delisting Risk Amplified	2020	link
R54	X (formally Twitter) user	Tweet by the handle @firoorg	2021	link
R55	Bitcoin News	Privacy-Centric Cryptocurrency Firo Suffers 51% Attack on Its Network	2021	link
R56	Cointelegraph	Privacy-focused Firo cryptocurrency suffers 51% attack	2021	link
R57	FXStreet	Bitcoin SV suspended on several crypto exchanges over risks of double spending attack	2021	link
R58	Cointelegraph	Breaking: BSV reportedly suffers 'massive' 51% attack	2021	link
R59	Forbes	Bitcoin Fork Suffers 'Massive' 51% Attack In Attempt To 'Destroy' The Cryptocurrency, Sending Its Price Sharply Lower	2021	link

E_8 – Ethereum Classic (attack period: 5-8 January 2019)

Early Tweet by an Expert (2 retweets):

2019-01-06 22:38:43+00:00
Unconfirmed Rumor: Alleged ETC double spend "75 blocks deep"
Ethereum Classic Network is operating normally.
According to #BlockScout blockchain explorer there have been no recent reorganizations.
See for yourself
blockscout.com/etc/mainnet/reâ€¦ https://t.co/MdYxfL8IbG

Tweet by an Expert (2 retweets, and then 25 retweets):

2019-01-07 06:55:30+00:00
"Chinese blockchain security firm SlowMist sent out an alert that the Ethereum Classic (ETC) network might have been targeted by a 51% attack."
Exclusive: One $ETC Private Pool Claimed over 51% Network Hashrate coinness.com/news/198264

Confirmation Tweet by etherchain.org (391 retweets)

2019-01-07 17:33:23+00:00
We can confirm that there was a successful 51% attack on the Ethereum Classic (#ETC) network with multiple 100+ block reorganization. We recommend all services to closely monitored the chain and significantly increase required confirmations.

Fig. 4. Example of expert Tweets for event E_8 of the first attack on Ethereum Classic [5-8 January 2019].

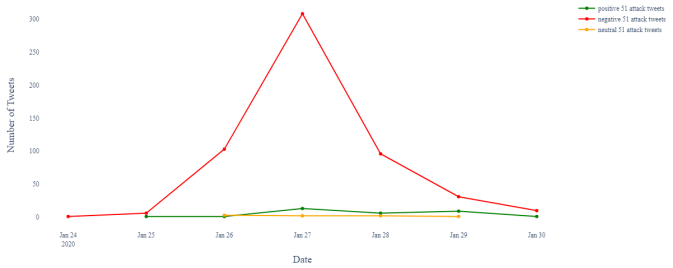


Fig. 5. Timeline of sentiments in $\mathcal{A}_{12}$ for event E_{12} of the second attack on Bitcoin Gold [23-24 January 2020]. The plot of the sentiment-wise volumes of tweets show that the peak was achieved on 27 January with over 300 negative tweets. Therefore, the peak day of this particular event is on the third day following the end date of the 51% attack.

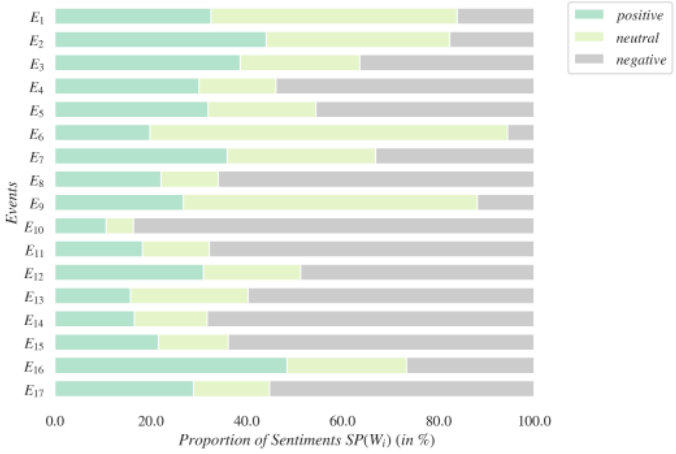


Fig. 6. $\texttt{SP}(\mathcal{W}_i)$: Sentiment profiles of the whole datasets $\mathcal{W}_i$

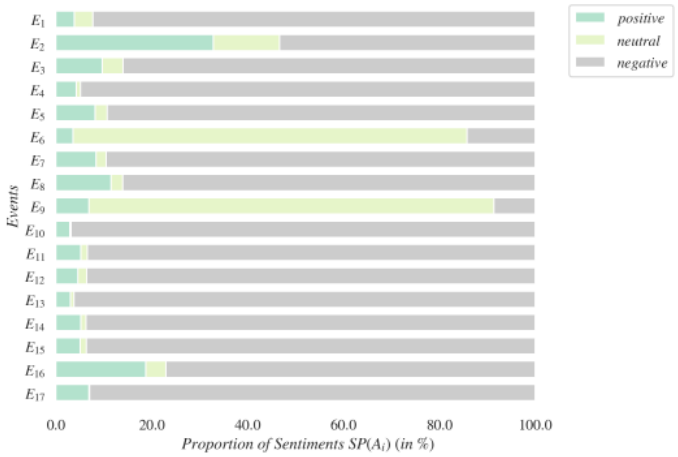


Fig. 7. $\texttt{SP}(\mathcal{A}_i)$: Sentiment profiles of attack datasets $\mathcal{A}_i$

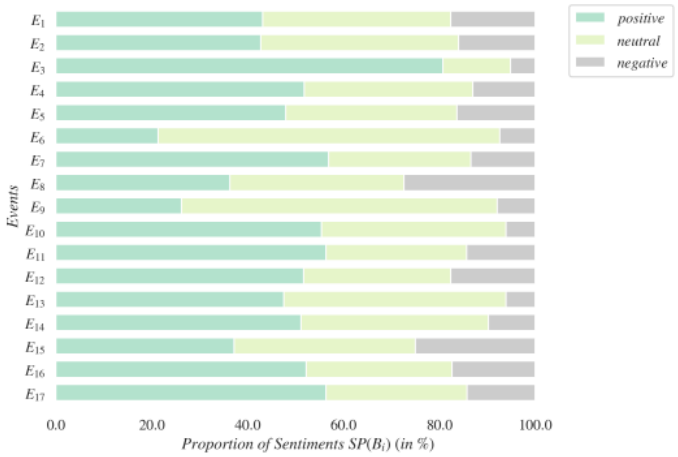


Fig. 8. $\texttt{SP}(\mathcal{B}_i)$: Sentiment profiles of the benchmark datasets $\mathcal{B}_i$

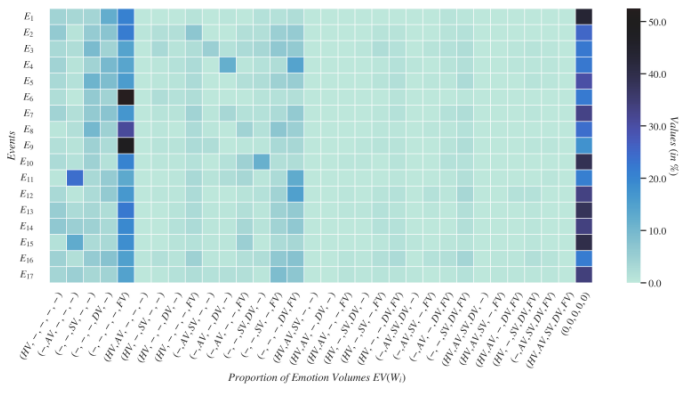


Fig. 9. Heat map of percentages of tweets carrying different emotions and their combinations in $\mathcal{W}_i$ datasets

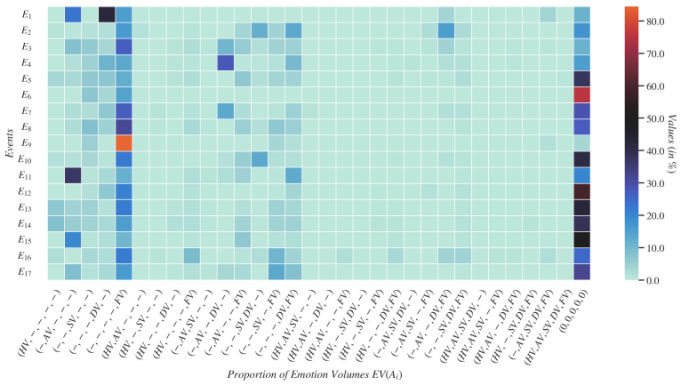


Fig. 10. Heat map of percentages of tweets carrying different emotions and their combinations in $\mathcal{A}_i$ datasets

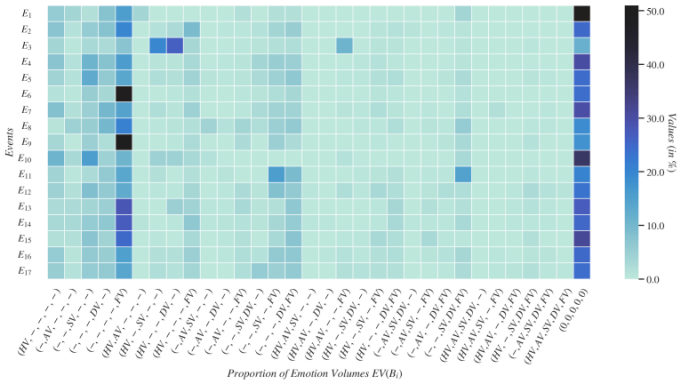


Fig. 11. Heat map of percentages of tweets carrying different emotions and their combinations in $\mathcal{B}_i$ datasets

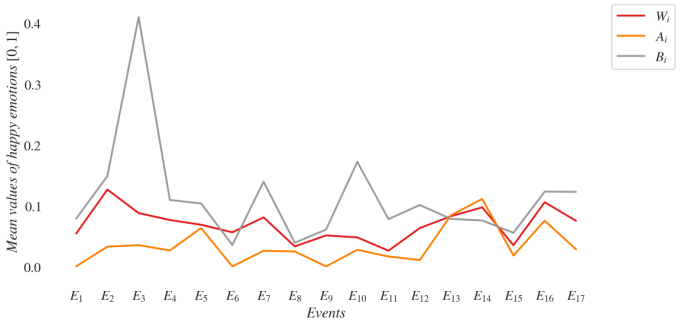


Fig. 12. $H(\mathcal{W}_i), H(\mathcal{A}_i), H(\mathcal{B}_i)$: Mean intensities of happiness across the datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$ for all events $E_i \in T$

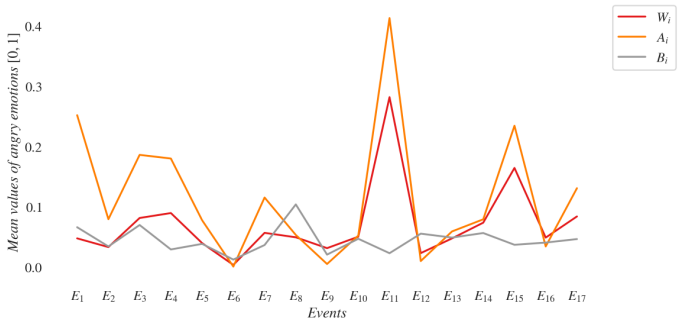


Fig. 13. $A(\mathcal{W}_i), A(\mathcal{A}_i), A(\mathcal{B}_i)$: Mean intensities of anger across the datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$ for all events $E_i \in T$

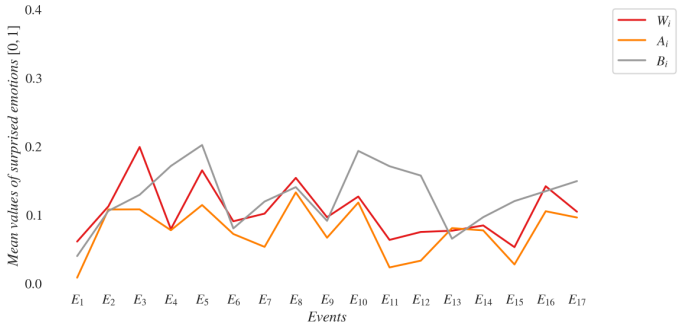


Fig. 14. $S(\mathcal{W}_i), S(\mathcal{A}_i), S(\mathcal{B}_i)$: Mean intensities of surprise across the datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$ for all events $E_i \in T$

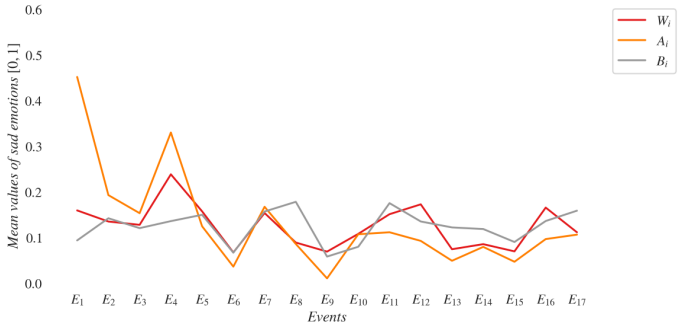


Fig. 15. $D(\mathcal{W}_i), D(\mathcal{A}_i), D(\mathcal{B}_i)$: Mean intensities of sadness across the datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$ for all events $E_i \in T$

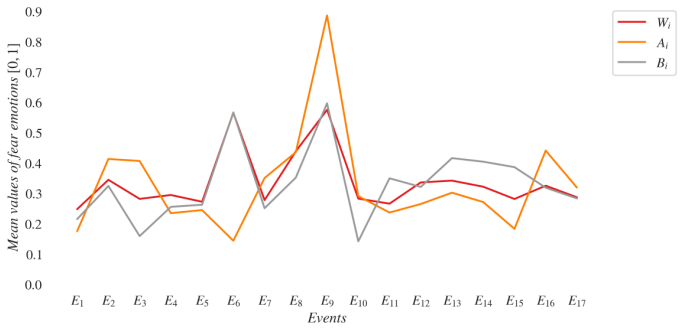


Fig. 16. $F(\mathcal{W}_i), F(\mathcal{A}_i), F(\mathcal{B}_i)$: Mean intensities of fear across the datasets $\mathcal{W}_i, \mathcal{A}_i, \mathcal{B}_i$ for all events $E_i \in T$