Kamps and Kleinberg *Crime Sci* (2018) 7:18
https://doi.org/10.1186/s40163-018-0093-5

Crime Science

RESEARCH Open Access

To the moon: defining and detecting cryptocurrency pump-and-dumps

Josh Kamps[1] and Bennett Kleinberg[2,3*]

Abstract

Pump-and-dump schemes are fraudulent price manipulations through the spread of misinformation and have been around in economic settings since at least the 1700s. With new technologies around cryptocurrency trading, the problem has intensified to a shorter time scale and broader scope. The scientific literature on cryptocurrency pump-and-dump schemes is scarce, and government regulation has not yet caught up, leaving cryptocurrencies particularly vulnerable to this type of market manipulation. This paper examines existing information on pump-and-dump schemes from classical economic literature, synthesises this with cryptocurrencies, and proposes criteria that can be used to define a cryptocurrency pump-and-dump. These pump-and-dump patterns exhibit anomalous behaviour; thus, techniques from anomaly detection research are utilised to locate points of anomalous trading activity in order to flag potential pump-and-dump activity. The findings suggest that there are some signals in the trading data that might help detect pump-and-dump schemes, and we demonstrate these in our detection system by examining several real-world cases. Moreover, we found that fraudulent activity clusters on specific cryptocurrency exchanges and coins. The approach, data, and findings of this paper might form a basis for further research into this emerging fraud problem and could ultimately inform crime prevention.

Keywords: Cryptocurrencies, Fraud, Pump-and-dump, Anomaly detection

Introduction

Cryptocurrencies have been increasingly gaining the attention of the public, and their use as an investment platform has been on the rise. These digital currencies facilitate payments in the online sector without the need for a central authority (e.g., a bank). The market for cryptocurrencies is rapidly expanding, and at the time of writing currently had a market capitalisation of around 300 billion US dollars (CoinMarketCap 2018) making it comparable to the GDP of Denmark (Cryptocurrency Prices 2018). Despite the vast amounts of money being invested and traded into cryptocurrencies, they are uncharted territory and are for a large part unregulated. The lack of regulation, combined with their technical complexity, makes them an attractive target for scammers who would seek to prey on the misinformed. One such scam is known as a pump-and-dump (P&D), where bad actors attempt to make a profit by spreading misinformation about a commodity (i.e., a specific cryptocurrency coin) to artificially raise the price (Kramer 2004). This scam has a long history in traditional economic settings, going as far back as London's South Sea Company in the 1700s (Brooker 1998), then found a natural home in penny stocks and on the Internet (Kramer 2004; Temple 2000), and has now recently appeared in cryptocurrency markets (Khan 2018; Mac and Lytvynenko 2018; Martineau 2018).

The academic literature on cryptocurrency (crypto) P&D schemes is scarce (for an exception, see the recent working paper of Li, Shin, & Wang, 2018). Thus, this paper will give an overview of what is currently known about the topic from blogs and news sites. To provide a theoretical angle, economic literature related to the topic is examined, and this information synthesised with cryptocurrencies by highlighting the similarities and potential differences. As these patterns are a type of anomaly,

*Correspondence: bennett.kleinberg@ucl.ac.uk
[2] Dawes Centre for Future Crime, Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK
Full list of author information is available at the end of the article

literature on anomaly detection algorithms is also discussed. The goal is to propose some defining criteria for what a crypto P&D is and to subsequently use this information to detect points in exchange data that match these criteria, forming a foundation for further research.

What is a pump-and-dump scheme?

A pump-and-dump scheme is a type of fraud in which the offenders accumulate a commodity over a period, then artificially inflate the price through means of spreading misinformation (pumping), before selling off what they bought to unsuspecting buyers at the higher price (dumping). Since the price was inflated artificially, the price usually drops, leaving buyers who bought on the strength of the false information at a loss. While we do not provide a rigorous crime script analysis (see Borrion 2013; Keatley 2018; Warren et al. 2017) here, Fig. 1 can be viewed as a script abstraction of three main stages—accumulation, pump, and dump. The accumulation phase usually occurs incrementally over a more extended period of time, in order to avoid raising the price before the pump.

What are cryptocurrencies?

Cryptocurrencies are a digital medium of exchange, and they usually rely on cryptography instead of a central institution to prevent problems like counterfeiting. For example, the most popular cryptocurrency is Bitcoin (BTC), and some of its benefits are that it allows for trustless and de-centralised transactions since it is impossible to reverse a payment, and there are no third parties (e.g., banks) involved (Nakamoto 2008). In traditional financial systems, a customer trusts the third-party (e.g., a bank) to update their ledger to reflect the customer's accounts balance. To the contrary, with Bitcoin, this ledger is distributed across a network, and everyone on the network possesses a copy and can—in principle—verify its contents. That public ledger is known as the blockchain and is the core technology upon which Bitcoin and many other cryptocurrencies rest. There are now many different types of cryptocurrencies, with less widely known ones referred to as 'altcoins', and they all run on slightly different technical principles, with different utilities and benefits (Bitcoin Magazine 2017). Besides Bitcoin, some of the other currently more popular cryptocurrencies include Ethereum (https://ethereum.org/), Ripple (https://litecoin.org/), and Litecoin (https://litecoin.org/).

Aims of this paper

In this paper, we set out to achieve three primary goals. First, absent a body of academic research on cryptocurrency pump-and-dump schemes, we provided an initial working formalisation of crypto P&Ds identifying criteria that might help in locating and ideally preventing this emerging fraud problem. Second, we utilise these indicators and propose an automated anomaly detection approach for locating suspicious transactions patterns. Third, to better understand the crypto P&D phenomenon, we zoom in on the exchange level and on the cryptocurrency pairings level. The overarching aim of this paper is to spark academic interest in the topic and to introduce P&Ds as an emerging problem.

Pump-and-dump schemes in the traditional economic context

In the early eighteenth century, con artists who owned stock in the South Sea Company began to make false

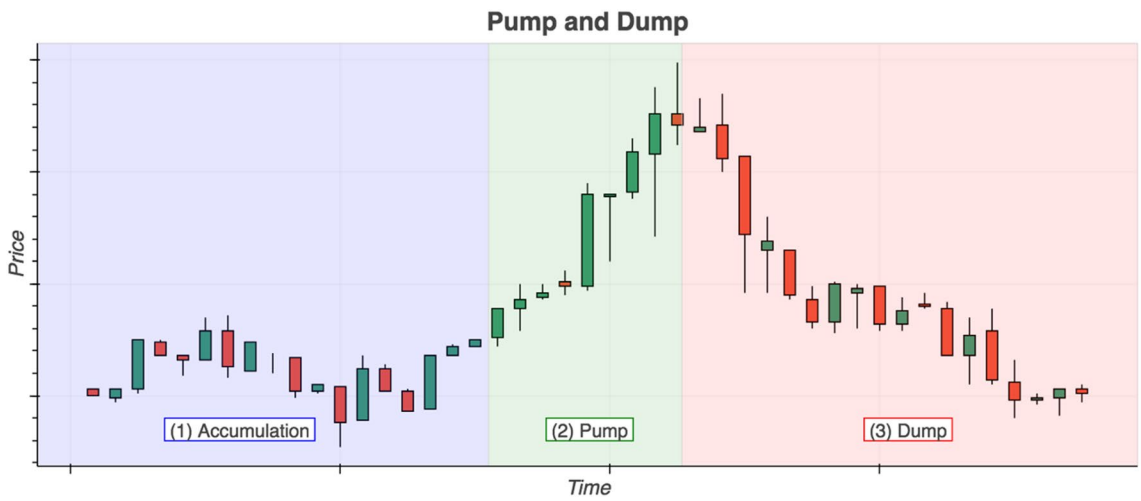


Fig. 1 Schematic abstraction of the three phases of a pump-and-dump operation

claims about the company and its profits. The goal was to artificially raise the price of the stock, and then sell it off to misinformed buyers who were led to believe that they were buying a promising commodity. This was referred to as the *South Sea Bubble* and serves as an early documented example of a P&D scheme (Bartels 2000; Brooker 1998).

In modern times, P&D schemes have predominantly been Internet-based focusing on so-called "penny" or "microcap" stocks, which are smaller companies that do not meet the requirements to be listed on the larger exchanges such as the NASDAQ (Dugan 2002; Temple 2000). Microcap stock exchanges are not held to the same standard of regulation, which implies that there is usually not as much information about the companies that are listed making them easier to manipulate. For example, in the US, large public companies file publicly available reports with the Security Exchange Commission (SEC) which are often analysed by professionals (US Securities and Exchange Commission 2017). Access to and the verification of information is typically more difficult with microcap companies. Misinformation about the stocks is often spread through email spam which has been found to have a net positive effect on the stock price (i.e., the spam is effective in increasing the price, see Bouraoui 2009). In the United States, it is illegal to run a P&D operation on penny stocks, and there are multiple cases of people having charges pressed against them for their participation in a P&D scam ("Developments in Banking and Financial Law: 2013," 2014; Yang and Worden 2015).

Pump-and-dump schemes in the cryptocurrency context

There is currently a lack of academic literature on cryptocurrency pump-and-dump schemes, so this section seeks to give an overview of the current landscape of cryptocurrency P&D schemes as they have been realised in various blog posts and news articles. In the cryptocurrency context there is an overall slightly different modus operandi than in the traditional context of penny stocks; specifically, this has been seen in the rise of dedicated public P&D groups. These groups have emerged in online chat rooms such as Discord (https://discordapp.com) and Telegram (https://telegram.org) with the sole purpose of organising pump-and-dump scams on select cryptocurrencies (Fig. 2). The number of members in some of these groups is reported to have been as high as 200,000, with smaller groups still running about 2000 (Martineau 2018). Price increases of up to 950% have been witnessed, demonstrating the extent of manipulation these groups are capable of (Thompson 2018). For these P&D groups to achieve the best results, several reports of activity show that they almost exclusively target less popular coins, specifically those with a low market cap and low circulation, since they are deemed easier to manipulate

Fig. 2 Example of a pump-and-dump chat group with over 40,000 members. Left: Telegram group 'Rocket dump'. Right: Corresponding exchange data (Binance) of the targeted coin (Yoyo) showing the effect of the pump. The yellow, purple, and maroon lines represent the moving average for the last 7, 25, and 99 days respectively

(Khan 2018; Mac and Lytvynenko 2018; Town 2018). Estimating the full scope of the damages caused by cryptocurrency pump-and-dumps is difficult; yet there is some evidence to show that such schemes are generating millions of dollars of trading activity. The Wall Street Journal published an investigative article that looked at public pump-and-dump groups and 6 months of trading activity. They found $825 million linked to pump-and-dump schemes, with one group alone accounting for $222 million in trades (Shifflett 2018). This gives a glimpse of how much monetary activity is generated by these groups, the impact of which could be even greater as many groups presumably operate in private or invite-only groups.

The pump-and-dump procedure usually consists of the group leaders declaring that a pump will take place at a particular time on a particular exchange, and only after the specified time will the coin be announced (see Fig. 2). After the coin is announced members of the group chat try to be amongst the first to buy the coin, in order to secure more profits. Indeed, if they are too slow, they may end up buying at the peak and be unable to sell for a profit. The 'hype' around buying the coin once the pump is announced is due to the short timescale of these schemes: Martineau (2018) reported on two pumps that reached their peaks within 5–10 min. During the pumping phase, users are often encouraged to spread misinformation about the coin, in an attempt to trick others into buying it, allowing them to sell easier. The misinformation varies, but some common tactics include false news stories, non-existent projects, fake partnerships, or fake celebrity endorsements (Martineau 2018; Town 2018). Consider the example where a group of offenders impersonated Internet entrepreneur John Mcafee's twitter account `@OfficialMcafee` by including an extra '`l`' in the username (Mac and Lytvynenko 2018). The fake account sent a positive tweet about a particular altcoin and all the users in the P&D group were told to retweet it. Within 5 min. The price of the coin had gone from $30,- to $45,-, collapsing back down to $30,- after about 20 min. Anything which creates a general air of positivity is fair game because the goal is to dump their coins on unwitting investors who have not done their due diligence, by preying on their fear of missing out on the next big crypto investment.

In a move to secure profit for themselves, many pump-and-dump group leaders will often use their insider information to their advantage: because they know which coin will be pumped, they can pre-purchase the coin for a lower price before they announce it. This guarantees them profit while leaving other users to essentially gamble on whether or not they can predict the peak. The fear of missing out and the potential to beat the odds might drive prospective cryptocurrency investors into joining a pump. Group leaders can also guarantee profits by offering access to the pump notification at an earlier stage prior to the group-wide announcement, in exchange for payment. Even a few seconds of temporal advantage are sufficient to potentially place buy orders before others, and thereby obtain cheaper coins, hence increasing the buyer's benefit from the of the pump-and-dump operation (Martineau 2018).

Due to the fact that the technology behind cryptocurrencies is relatively new, and that most exchanges are unregulated, pump-and-dump manipulation is currently not always illegal; and even where it is, it cannot always be easily enforced. However, governing bodies are beginning to realise the problem, and in the United States the Commodity Futures Trading Commission has issued guidelines on how to avoid P&D scams, as well as offering a whistle blower program (U.S. Commodity Futures Trading Commission 2018).

Defining a cryptocurrency pump-and-dump

Mitigating and preventing pump-and-dump schemes will require knowledge about their operation, and thus the detection of these pump-and-dump schemes is a step towards the goal of mitigation. To begin searching for and identifying potential P&D type patterns in exchange data, a working definition for what constitutes a P&D is needed. A proposal for defining criteria will be given in this section by summarising the insights regarding traditional and crypto P&D schemes that have been outlined in the previous section. Table 1 summarises some of the key similarities and differences with the respect to the target, tactic, and timescale of traditional penny stock and crypto pump-and-dump schemes.

Table 1 indicates that a crypto P&D seems similar to a penny stock P&D in that assets that share the same properties are targeted. However, in general, it appears that as a result of different tactics the time scale has been narrowed and moved towards near real-time. Just as the digitisation of information via the Internet increased the

Table 1 Comparison of traditional and crypto pump-and-dump schemes

	Traditional	Crypto
Target	Low market cap	Low market cap
	Low volume	Low volume
	Low price	Low price
	Lack of reliable information	Lack of reliable information
Tactic	Misinformation	Real-time misinformation
	Privately organised (smaller scale)	Public or private group scams (larger scale)
Timescale	Medium (days to weeks)	Short (minutes to hours)

rate of P&D scams on penny stocks, so too it seems the digitisation of currency itself has increased the rate and speed at which a P&D can take place.

Using the identified characteristics of crypto P&Ds allows us to formulate criteria that could be helpful in detecting P&D patterns in exchange data (Table 2). Specifically, we argue that indicators of P&Ds can be subdivided into *breakout indicators* which refer to the signals that will always be present during a pump-and-dump, and *reinforcers* which refer to indicators which may help increase the confidence that the observed data point is the result of manipulation. The volume and price are discussed with an *estimation window*, referring to a collection of previous data points, of some user-specified length. For example, a moving average over a previously defined time period could be used, which would allow for discussing spikes with regards to some local history. This is not to say that the proposed criteria are sufficient to encompass all crypto P&Ds. Instead, we chose to resort to conservative criteria that are necessary for a P&D and that appear to have emerged based on the information in the previous section.

Table 2 Indicators of pump-and-dumps per temporal dimension and indicator type

	Temporal dimension	
	Real time indicators	**Post-pump indicators**
Breakout indicators		
Volume	Has the volume at the current data point been significantly higher than in the estimation window?	Was there a decline in volume after the event window where a pump was detected?
Price	Has the price at the current data point been significantly higher than in the estimation window?	Was there a decline in price after the event window where a pump was detected?
Reinforcers		
Market cap	Is the market cap of the coin relatively low? (+)	
Number of exchanges	Whether the coin is listed on multiple exchanges and the indicators only spike on one (+) Whether the coin is listed on multiple exchanges and the indicators spike on multiple exchanges (neutral) Whether the coin is not listed on multiple exchanges (+)	
Symbol pair	Whether the coin is trading for BTC or some other cryptocurrency (+) Whether the coin is trading for USD or some other fiat currency (−)	

A (+) denotes an increase in confidence for a pump, while a (−) denotes a decrease in confidence. A symbol pair is a term used to denote which currency is trading for which, thus BTC/USD is a symbol pair representing that Bitcoin is being traded for US dollars

Method

Data

To obtain data for analysis, the CCXT (Ccxt 2018) library was used which provides a unified way to programmatically access the data from a variety of cryptocurrency exchanges using the python programming language. Despite the unified access, the exchanges still differ in the amount of historical data they serve, and in the cryptocurrencies, they have listed. Therefore, decisions had to be made on what data to obtain.

Data availability statement

The data and code to reproduce the analysis and data retrieval are publicly available at https://osf.io/827wd/.

Format of cryptocurrency exchange data

Cryptocurrencies are listed on exchanges in symbol pairs denoting which currencies are trading for which. For example, to trade Litecoin (LTC) for Bitcoin (BTC), the symbol pair listed is "LTC/BTC". Exchange data are returned as a set of Open High Low Close Volume (OHLCV) entries, detailing the trading data for that particular moment in time. Table 3 shows an example of the OHLCV terminology in its raw representation and Fig. 3 shows the candlestick chart representation of OHLCV data. The top and bottom wicks represent the highest and lowest value respectively, while the coloured candle represents whether the closing price was higher than the opening price (green) or lower than the opening price (red). The top of a green candle is the closing price, and the bottom is the opening price, and vice versa for a red candle. Candles can represent a variety of timeframes, but they often represent 30 min, 1 h, or 24 h. Smaller candle sizes mean more data per time period, so usually the smaller the candle size, the fewer days one can retrieve from an exchange, due to imposed limitations on the amount of data retrievable using their API. One-hour candles were chosen as a compromise between the resolution of the data and the amount of historical data available.

Obtaining the data

The CCXT library (https://github.com/ccxt/ccxt) supports access to 115 different cryptocurrency exchanges. However, not all of these permit the public retrieval

Table 3 An example row of OHLCV data

Timestamp	Price				Trading volume
	Open	High	Low	Close	
2018-04-20 01:00:00	0.11804	0.11882	0.11758	0.11881	181.16102255

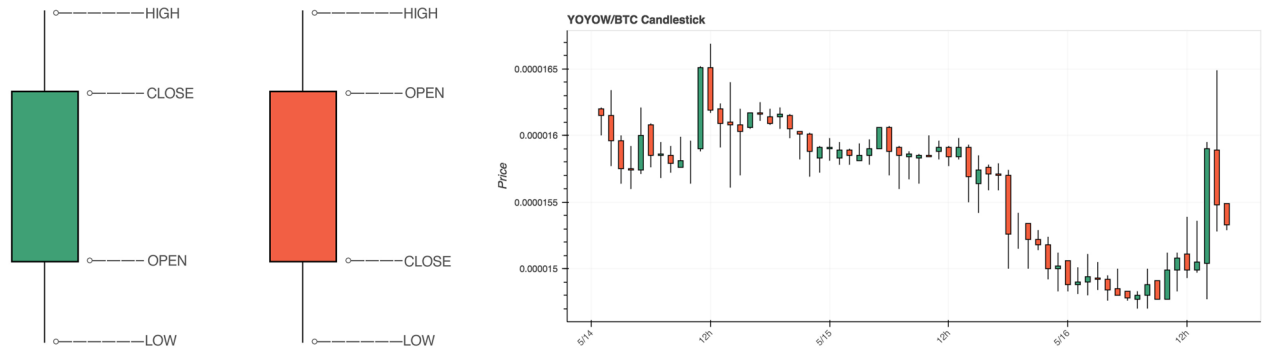


Fig. 3 An example of a candlestick chart for the YOYOW/BTC trading pair

of historical data. After filtering for those conditions, 24 exchanges remained. To make the results more robust, the 24 candidate exchanges were filtered further to exchanges with at least 50 symbol pairs and at least 20 days of historical 1-h OHLCV data. In total, five exchanges matched all the criteria, and 480 candles (~20 days) of data for every available symbol pair were pulled from each of these exchanges (see Appendix).

Analytical approach

A successful P&D will often exhibit a marked spike in price and volume (see Table 2) that can easily be detected by human observation. However, with hundreds of exchanges and symbol pairings, and trading transactions not bound to specific times during the day, it is impractical and infeasible to resort to a manual approach for the detection of P&Ds only. Therefore, we resorted to an automated detection approach using anomaly detection.

A brief introduction to anomaly detection

Data points which do not conform to the rest of a dataset are often referred to as anomalies or outliers. Anomaly detection is the process of identifying these non-conforming points (Chandola et al. 2009). Anomaly detection techniques can be broadly categorised into supervised and unsupervised anomaly detection. Supervised anomaly detection relies on a training data set to learn what "normal" is for the domain. The latter hinges on the ability to acquire an adequately sized training set, something which is often challenging. Conversely, unsupervised techniques rely on the assumption that anomalies are a rare occurrence in the data to prevent an excess of false signals. Here, it is the researcher's or analyst's task to determine the parameters that constitute an anomaly.

Types of anomalies

There are various types of anomalies, which have been grouped into three major categories by Chandola et al. (2009): point anomalies, collective anomalies, and contextual anomalies. Point anomalies are merely points in the data which are anomalous to the rest of the data. An example would be an unusually large purchase relative to an individual's historic spending behavior. Collective anomalies, on the other hand, refer to a situation in which one single data point may not be anomalous by itself. Instead, a co-occurrence or temporal proximity of anomalous data points might indicate behavior that is anomalous (e.g., a human electrocardiogram in which a single low point would not necessarily be anomalous, but consecutive low values would be indicative of a problem). Finally, contextual anomalies (also known as 'conditional anomalies', Song et al. 2007) are data points which would only be considered anomalous in specific contexts. For example, a warm temperature in the winter would be anomalous, but in the summer would be considered normal.

Anomaly detection in the context of crypto P&D schemes

In the context of this paper, unsupervised anomaly detection will be the focus, as no labelled training data is currently available for cryptocurrency pump-and-dump schemes (see Discussion). Conditional anomalies consider contextual information about the setting (Song et al. 2007). This is described through *indicator variables*, of which the values may be directly indicative of an anomaly, and *environment variables*, whose variables are not directly indicative of an anomaly. The indicator variables are determined to be anomalous depending on the values of the environmental variables. In the current context this means the goal is to locate the breakout indicators, with respect to the reinforcers (Table 2). For the scope of this paper, we do not consider the reinforcer of whether a symbol pair was present on multiple exchanges, due to the amount of data available. Thus, the goal is to locate corresponding price and volume spikes of coins with a low market cap that are trading for other cryptocurrencies. Due to the nature of P&D schemes, pumps are

inherently local phenomena, so the goal is to detect local anomalies concerning recent history (i.e., to detect *local conditional point anomalies*).

Anomaly anatomy

The anomaly detection technique utilised is a thresholding technique, inspired by previous research regarding denial of service attacks on a network (Siris and Papagalou 2004). For a particular value, a simple moving average is computed by taking the average of previous values in a given time window, the length which is known as the lag factor. In this way, one can compare a value to the trend over a time period, as opposed to a singular value, allowing for the detection of local anomalies in comparison to recent history. This type of thresholding algorithm, allows us to provide a functioning baseline which further research could then expand upon with more sophisticated algorithms. Additionally, as more is learned about cryptocurrency pump-and-dump schemes, it is likely that more domain information (e.g., certain times, coins, or trading patterns) can be incorporated into the algorithms in an effort to increase the detection accuracy.

Price anomaly

If the high price at any given point is greater than the computed anomaly threshold for that point, then the point is determined to be anomalous. The anomaly threshold is computed using a given percentage increase ϵ, a lag factor γ and the simple moving average $\mu_\gamma(x)$ over the closing price. An instance x is a particular observation in the time series that is associated with the respective OHLCV values. In this case, x and γ can be considered as datetime objects, therefore $x - \gamma$ would indicate moving backwards in the time series by a factor of γ. The moving average is thus $\mu_\gamma(x) = \frac{\sum_{i=x-\gamma}^{x} x_{close}}{\gamma}$

which is defined for all x where $x - \gamma \geq 0$. The threshold for any given point after the time lag is defined as $\epsilon \cdot \mu_\gamma(x)$ giving us the point anomaly function:

$$price\_anomaly(x) = \begin{cases} True, x_{high} > \epsilon \cdot \mu(x) \\ False, x_{high} \leq \epsilon \cdot \mu(x) \end{cases} \quad (1)$$

Volume anomaly

The volume anomaly is defined almost identically to the above, except with the moving average computed as $\mu_\gamma(x) = \frac{\sum_{i=x-\gamma}^{x} x_{volume}}{\gamma}$, resulting in:

$$volume\_anomaly(x) = \begin{cases} True, x_{volume} > \epsilon \cdot \mu(x) \\ False, x_{volume} \leq \epsilon \cdot \mu(x) \end{cases} \quad (2)$$

Pump anomaly

The goal is to detect local conditional point anomalies, that is the co-occurrence of both a price anomaly and a volume anomaly. Additionally, the contextual information of whether or not the coin has a low market cap or is a crypto/crypto trading pair can be considered. There are perhaps other contextual indicators that could be investigated, though for the scope of this paper, only the two mentioned above will be looked at.

Low market cap

The market cap of a coin is defined as its price times the supply, and represents a way of judging the popularity, or size, of a coin. The market cap data were pulled from https://coinmarketcap.com/. The top ten coins from the dataset and the percent of the total market cap they account for are shown in Table 4. From this it can be seen that the top ten coins account for over 85% of the total market capitalisation, implying that a vast majority of coins have a much smaller market cap relative to the top. For the rest of this paper, "low market cap" will be defined as any coin below the 75th percentile (0.029%) of the total market cap.

Results

This section investigates various values for the different parameters and shows how changing these affects the results found, with the goal of providing a suggestion for balanced parameters. Hopefully, these parameters could then be taken to a real-time system, to be further monitored and tuned as time progresses.

Locating crypto pump-and-dumps

It is possible to formulate expectations based on the domain information presented in earlier sections. Since low market cap coins are targeted more often, we would expect to see more P&Ds amongst that group of coins.

Table 4 The top 10 coins by percentage of market cap

Coin	% of total market cap
BTC	42.0
ETH	19.4
XRP	7.8
BCH	5.6
EOS	3.5
LTC	2.1
XLM	1.6
ADA	1.6
TRX	1.2
USDT	0.95

Similarly, crypto/crypto symbol pairs would also be expected to exhibit more P&D activity. Additionally, since this paper only simulates real-time detection, it is possible to look forward in time, and see which of the alleged pumps were followed by a marked drop in price, which could be an indication of users dumping their coins, making it more likely that the preceding pump was the result of nefarious activity (i.e., a pump-and-dump).

Anomaly detection

Initial parameters

The idea behind the initial parameters for the detection system was to start off relatively 'weak', to give an initial starting point. We chose a 12 h estimation window, 25% volume increase and a 3% price increase. The results show that the 25% volume increase threshold was perhaps too low, due to the abundance of volume spikes found. Similarly, the 3% increase threshold for the price spikes also proved to be a bit too low, as indicated by (Table 5). This led to finding over 9000 alleged pump-and-dumps across the dataset, which is an average of about nine P&Ds per coin over 20 days. While these may be interesting points to investigate, making the parameters stricter could help reduce false positives (i.e., false flags). Ultimately the goal is to find a set of balanced parameters that filter the points detected down to a more reasonable number that can then be further assessed by humans. The percentage of spikes that were found to have corresponding price dips was quite high with the initial parameters (90%), but this could be due to the vast number of spikes detected, to begin with. Figure 4 shows an example of an annotated candlestick chart using the initial parameters.

Strict parameters

We increased the estimation window to 24 h, so it required a more drastic change in comparison to the average. Additionally, the volume and price thresholds were increased to 400% and 10% respectively (Fig. 5). This led to detecting 920 alleged pump-and-dumps over 20 days, about 0.5 P&Ds per symbol. Price dips followed only 50% of the alleged pumps, and the total number of pump-and-dumps was consequently lower than with the initial parameter set.

Balanced parameters

With the information gained from the previous two parameter sets, we attempted to find a balance between the two. The estimation window was returned to 12 h to constrain the search locally, and the volume and price thresholds were a compromise between the initial and strict parameter values, at 300% and 5% respectively. This resulted in about 1.6 pump-and-dumps per symbol, for a total of 2150 over the 20 days of data (Fig. 6). Moreover, 75% of the alleged pumps were found to have corresponding price dumps; which could mean that in a real-time system, these parameters could lead to detecting points that would often be flagged for further investigation because they are possibly indicative of a P&D scheme.

Closer inspection of the balanced parameter set

The results of the balanced parameter set were investigated closer to identify P&D dynamics at the exchange- and symbol pair-level. To do so, we filtered the results to only include observations where the P&Ds detected were on crypto/crypto symbol pairs with a low market cap.

Table 5 Results of the anomaly detection for three different parameter sets

	Initial parameters	Strict parameters	Balanced parameters
# of alleged pumps	9668	920	2150
# of pump and dumps	8738	485	1617
% P&D	90.4%	52.7%	75.2%
P&D/symbol	8.94	0.50	1.66
Crypto/crypto pair P&D %	96.1%	97.9%	97%
Low market cap P&D %	77.5%	84.9%	81.76%
Parameter estimation window	12 h	24 h	12 h
Parameter volume increase	25%	400%	300%
Parameter price increase	3%	10%	5%
Parameter price drop	Rolling average + 1 SD	Rolling average + 1 SD	Rolling average + 1 SD

Alleged pumps = pumps detected in real-time without taking information about whether the price drops afterwards; pump-and-dumps = those which are followed by a price dip of the rolling average of the previous observation plus one standard deviation; % of P&D = the percentage of the alleged pumps which were actually followed by price dips; % of crypto/crypto pairs = the percentage of all P&Ds which are made up of crypto/crypto trading pairs; % of low market cap P&Ds = the percentage of all P&Ds which are made up of low market cap coins

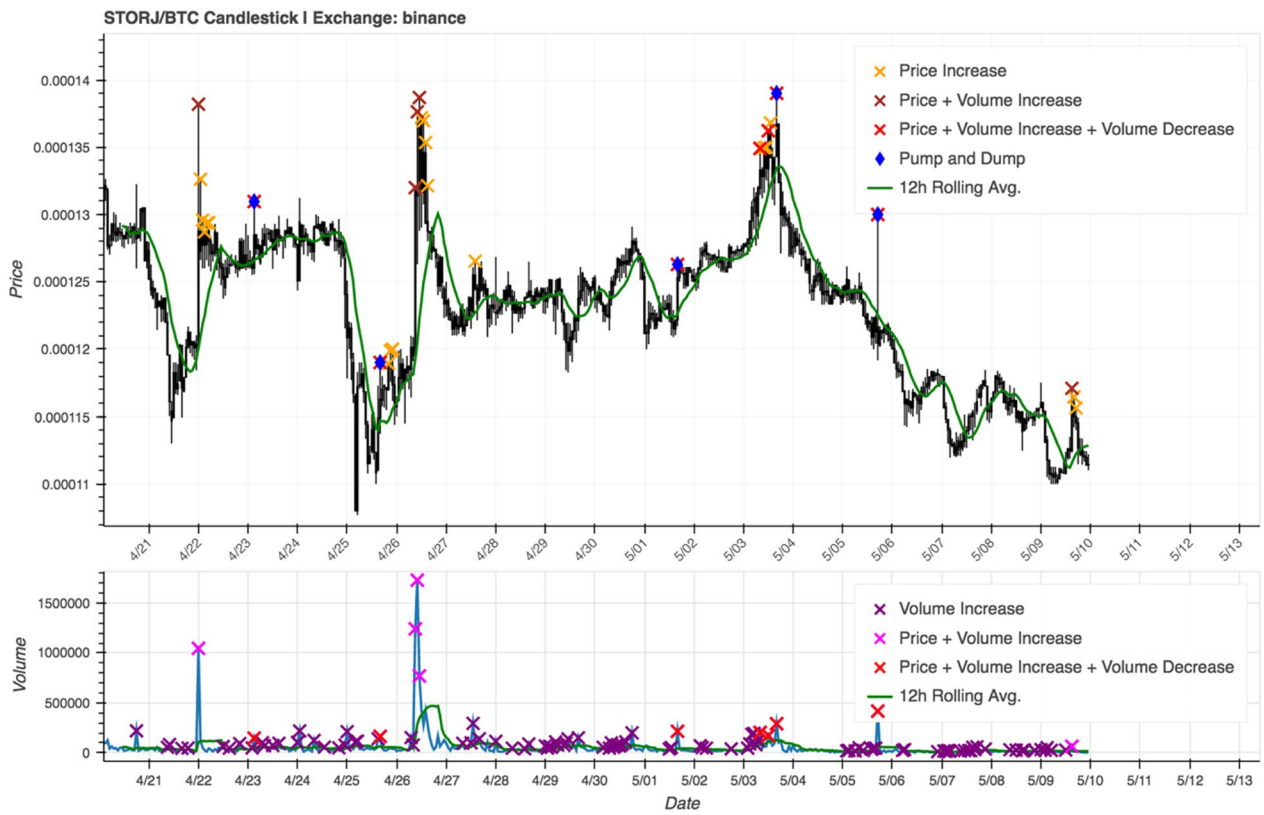


Fig. 4 Candlestick chart with anomaly detection indicators for the initial parameter set

Exchange-level findings

The number of P&Ds can be investigated on an exchange level, offering insight into which exchanges may be suitable targets for further investigation and mitigation techniques. An illustration of how the percentage of symbols analysed relates to the percentage of pumps detected is shown in Fig. 7. The exchanges *Binance* and *Bittrex* account for more of the pumps than the relative number of symbols analysed, suggesting these exchanges are utilised more for P&D schemes than others. Conversely, the exchange *Kraken* accounts for almost 6% of the symbols, yet less than 1% of the pumps. This is perhaps best explained by the fact that Kraken is one of the more regulated US-based exchanges, and deals mainly with crypto/fiat currency pairs, as opposed to crypto/crypto. These findings suggest that exchanges which offer more regulated trading would be less susceptible to P&D schemes.

Symbol pair-level findings

Breaking down the pump-and-dumps on a symbol level allows for a look into which cryptocurrencies, are disproportionately often affected, and hence more vulnerable (Table 6). The data show that the most P&Ds for one symbol pair was 13, with the vast majority of symbols having between 0 and 3 P&Ds. This is consistent with the notion that specific coins may be targeted more often than others. Also interesting to note is that five of the top ten most pumped coins were pumped on the *Bittrex* exchange. Further research could perhaps investigate the properties of these coins, in an attempt to see if there are links between the most pumped coins.

Figure 8 shows almost 9 days of candlestick data for the coin with the most P&D patterns detected. The individual spikes have been muted in the figure, to highlight only the pump-and-dumps. The resulting graph depicts rather suspicious trading activity, with many periods of lower price and volume, followed by significant spikes in both. During the 9-day period shown eight pumps were detected. This type of trading activity would be consistent with the activity of P&D groups organising multiple attacks on a single vulnerable coin. Regardless of whether it is directly the result of nefarious activity, it is still a pattern which raises question.

Real-world detectability

A core test of a pump-and-dump identification system is its real-world detectability. We used pump-and-dump schemes that we were explicitly orchestrated in online

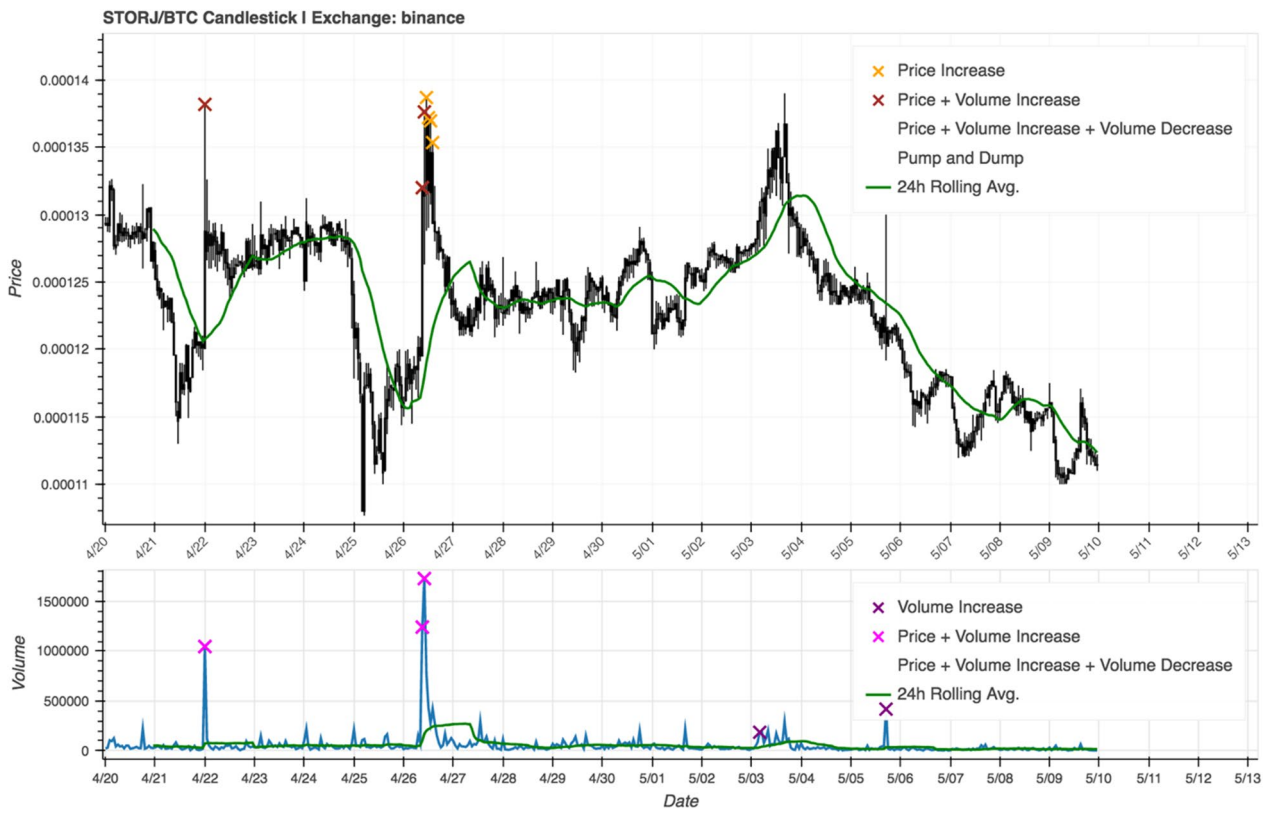


Fig. 5 Candlestick chart with anomaly detection indicators for the strict parameter set

chat groups as the 'gold standard' of confirmed cases. Albeit to a smaller extent, this source of confirmed P&Ds allows us to look at the detectability on a case-wise basis. The confirmed P&Ds were obtained by monitoring two pump-and-dump groups, *Moonlight Signal* (ca. 3000 members) and *Crypto Trading*™ (ca. 56,000 members) and observing their announcements. Using this information, we illustrate two cases where our system (with the balanced parameter set) successfully detected a confirmed P&D, and two cases where our system could not clearly identify the P&D.

Successful detection

Case 1 In Case 1 (Fig. 9) the coin that was to be victimised was announced on the 17th of August 2018, at 4 p.m. As a result of their coordinated efforts a large price and volume spike is visible, beginning exactly at the time at which the announcement took place. Our system was able detect the anomalous spikes, and correctly flagged the strange trading activity as being the result of a P&D.

Case 2 The announcement time for the P&D in Case 2 (Fig. 10) was the 21st of August 2018, at 4 p.m. Once again, the warning signals of corresponding price and volume spikes are present, and the system correctly marks the strange activity at the announced starting time as fraudulent. In this case we also observe the price and volume beginning to increase just prior to the announcement time, perhaps indicating insider trading by the group leaders.

Unsuccessful detection

Case 3 The pump announcement in this case was given on the 4th of September 2018, at 3:30 p.m. Once again, we observe corresponding price and volume spikes (Fig. 11), yet in this case our system failed to mark them as being

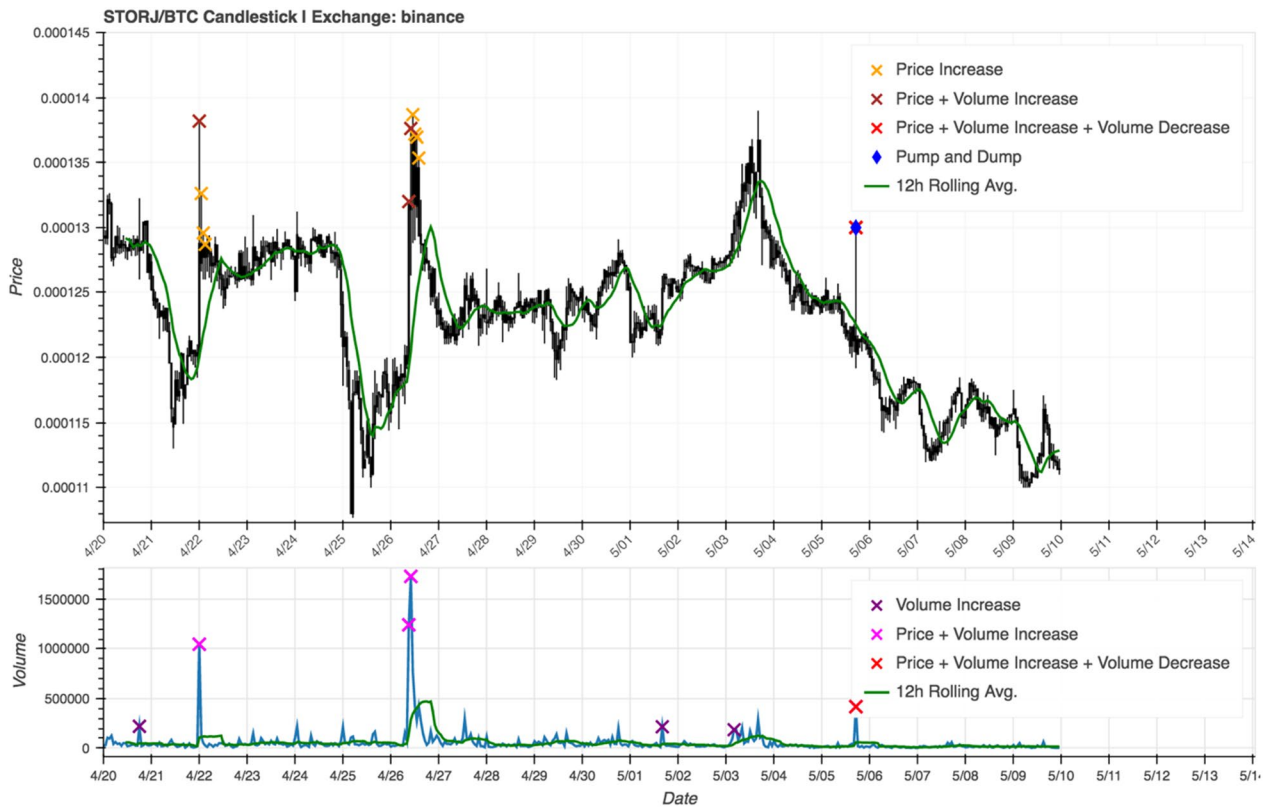


Fig. 6 Candlestick chart with anomaly detection indicators for the balanced parameter set

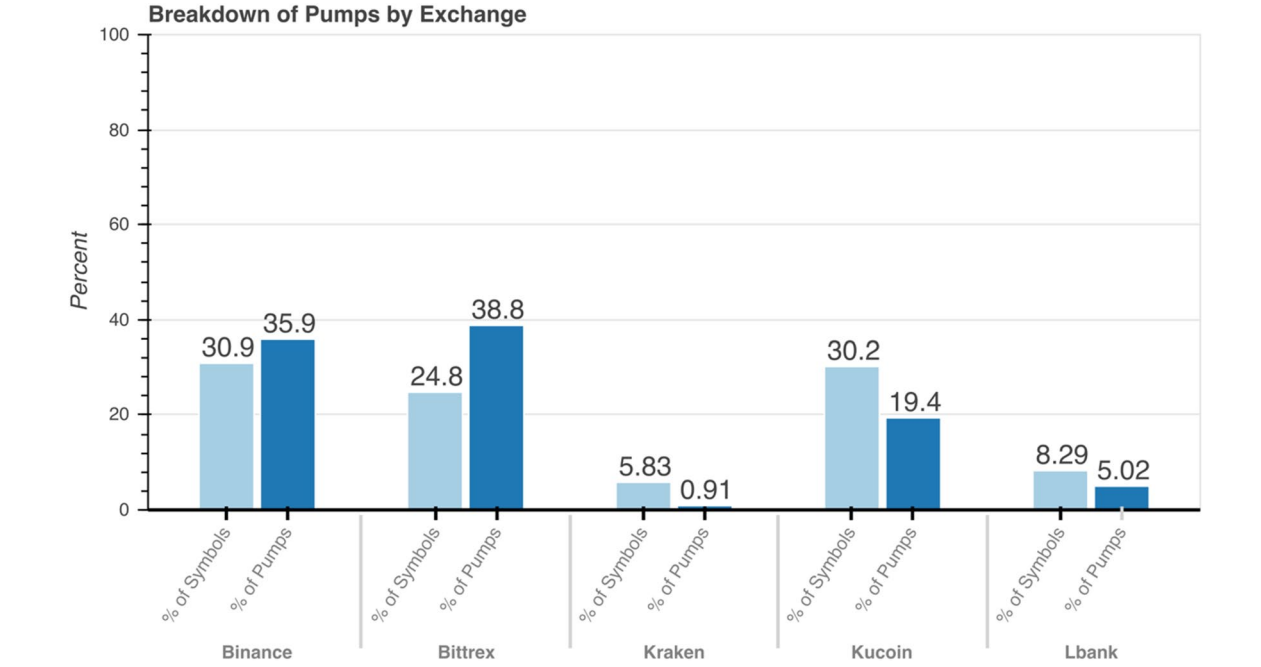


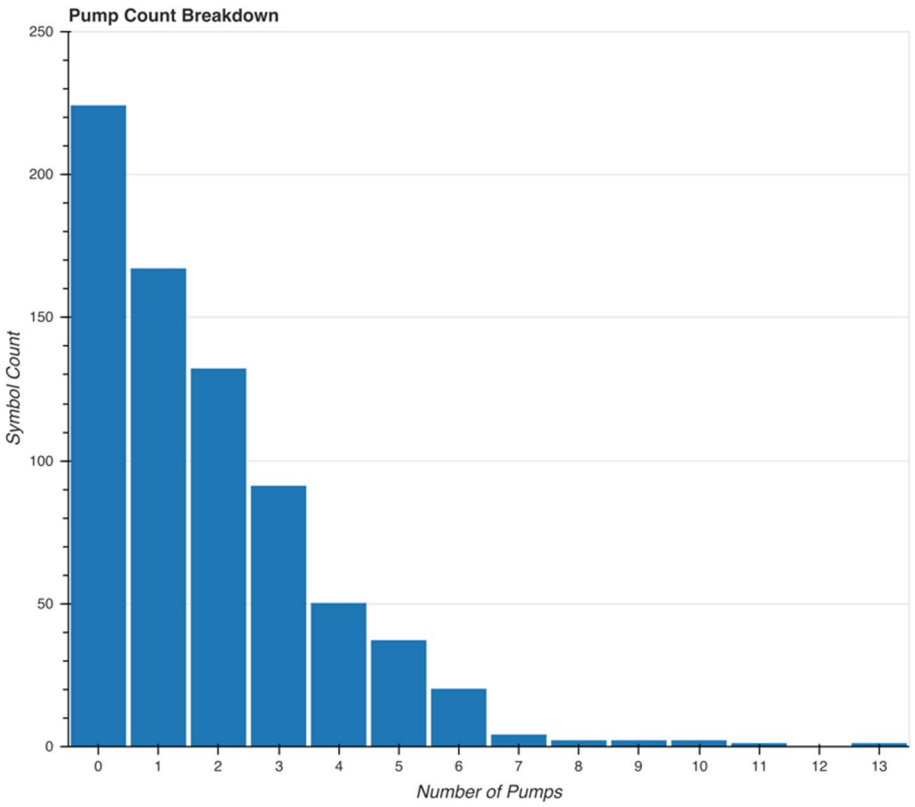
Fig. 7 The percentage of symbols and alleged pumps per cryptocurrency exchange

Table 6 Findings for the symbol pair-level inspection of alleged pump-and-dumps

Top 10 symbols ranked by # of pump-and-dumps

Exchange	*Symbol pair*	*Pump and dumps*
Lbank	DBC/NEO	13
Kucoin	CAPP/BTC	11
Lbank	TKY/ETH	10
Bittrex	DCT/BTC	10
Bittrex	BRX/BTC	9
Binance	MDA/ETH	9
Bittrex	RBY/BTC	8
Bittrex	EMC/BTC	8
Kucoin	ADB/BTC	7
Bittrex	GNT/ETH	7

Distribution of pump-and-dumps



the result of a pump-and-dump scheme. The reason for this is that the price continued to climb for a while after the pump, instead of immediately dumping. Thus, we can observe that sometimes the momentum caused by a pump group may actually persist for a period of time (in this case about 24 h). The coin being pumped in this case (RDN) was also pumped by the same group about 13 days previously (see "Case 2"); lending support to the idea that certain coins are targeted more often than others.

Case 4 In Case 4 (Fig. 12) the pump announcement was made at 4 p.m. on the 3rd of September 2018. Similarly, to Case 3, our system again fails to mark the anomalous spikes as a pump-and-dump, for the same reason of the price not dipping quickly enough afterwards. In order to correctly identify these cases in which the price maintains momentum for some time after the announcement, a potential improvement could be made to the algorithm whereby decreasing volume is also taken into consideration. That way, if either the price, or the volume dips, it is counted as a P&D, as opposed to only relying on price dips. Additionally, in this case, we see that the following day a P&D is detected by our system, though it is unknown whether this is a result of additional targeting by the group, or merely a false positive.

Discussion

This paper attempted to introduce to the crime science community the problem of cryptocurrency pump-and-dump schemes. With cryptocurrencies becoming increasingly popular, they are also becoming a more likely target for criminal activity. Cryptocurrency pump-and-dump schemes are orchestrated attempts to inflate the price of a cryptocurrency artificially. We identified breakout indicators and reinforcers as criteria for locating a pump-and-dump and investigated the data using an anomaly detection approach. While the choice of parameters that define an anomaly is inherently subjective, we

Fig. 8 A candlestick chart of the most pumped coin

observed that a balanced approach between the naïve initial parameters and the strict parameters might help in flagging suspicious trading activity. We were also able to show that using a limited set of parameters it is possible to detect *pumping* activity in the data as well as subsequent *dumping* activity. Moreover, we monitored two pump-and-dump groups in order to obtain several cases of real life pump-and-dump schemes which we then applied our detection algorithm to, in order to demonstrate its performance in real scenarios.

Pump-and-dumps as a challenge for crime science

Besides locating potential pump-and-dumps, we found evidence of clustering in the data. The vast majority of the coins are ones with a low market cap while the top ten coins accounted for 85% of the market cap. Furthermore, the final distribution of the pump-and-dumps showed that about 30% of the symbols accounted for roughly 80% of the pumps, indicating that even amongst low market cap coins, some coins are targeted more frequently than others. Translated to the environmental criminology literature, this pattern resembles repeat victimisation (Farrell and Pease 1993; Kleemans 2001; Weisel 2005; Farrell 2015). If a P&D chat group, for example, finds a suitable coin that they targeted successfully before, it is possible they may be more likely to perform another pump on that same coin; an example of this was shown in the case study section, where the group *Moonlight Signal* targeted the same coin (RDN) twice, in about a 2-week period. The clustering can be exploited for preventative purposes since efforts can be concentrated towards the clusters, finding out what makes them attractive targets, and implementing strategies to help mitigate potentially nefarious activity. Ideas from situational crime prevention, for example, such as increasing the risk or effort required to commit a P&D could also serve as useful methods for prevention (Clarke 2012). Consider an exchange which requires additional verification for users trading certain symbol pairs which are determined to be vulnerable. Such an intervention would increase the effort required to trade and hence to pump the vulnerable coin. When considering how to increase the risk, an example could be a system in which the automated detection of anomalous trading activity is used in cooperation with humans. That system could mark suspicious points which observers may then investigate further, increasing the chances that such P&D schemes are detected.

Fig. 9 The chart depicts the results of a pump-and-dump promoted by the group *Moonlight Signal*, which was signalled to commence at 4 pm (UTC) on the 17th of August. Anomalous price and volume spikes at the specified time are clearly visible, and the suspicious activity was correctly marked as a P&D scheme by our detection system. *Symbol:* OAX/BTC. *Exchange:* Binance

A major challenge for pump-and-dump prevention might lie in coordinating the efforts between private bodies such as cryptocurrency exchanges and government bodies. While governments are catching up on the problem and have allocated more resources to the mitigation of pump-and-dump schemes, exchanges might have little incentive to cooperate because they benefit from trading activity on their platforms. Finally, a move towards more government regulation—in our data less regulated exchanges were targeted disproportionately more frequently—might undermine the very concept of cryptocurrency trading as a decentralised exchange without government interference. An interdisciplinary, problem-oriented approach from both the practitioners' and the research community seems a path worthwhile exploring in the mitigation of cryptocurrency pump-and-dump schemes.

Limitations

In the current investigation, we resorted to publicly available data and provided a framework for the future analysis of cryptocurrency pump-and-dumps. However, several limitations merit attention. First, the accuracy of flagging an alleged pump-and-dump is dependent upon the parameters chosen and cannot be ascertained absent a ground truth of confirmed pump-and-dumps. Our analysis should be treated as a first attempt to place the topic in the academic literature. Second, the dataset only covers 20 days of data with hourly granularity. While this was sufficient for the scope of this paper, future research

Fig. 10 The chart depicts the results of a pump-and-dump promoted by the group *Moonlight Signal*, which was signalled to commence at 4 pm (UTC) on the 21st of August. Anomalous price and volume spikes at the specified time are clearly visible, and the suspicious activity was correctly marked as a P&D scheme by our detection system. *Symbol:* RDN/BTC. *Exchange:* Binance

would want to attempt to collect more substantial quantities of data and at a smaller granularity (e.g., per minute). Third, as with any flagging system, there is a decision to be made how many false positives are acceptable (i.e., incorrectly flagged coins). Arguably, an exchange would want to avoid announcing a coin of being used for fraudulent activity if this were not the case. This compromise is particularly complex in real-time settings so an interesting alternative avenue for future research might be to move towards the identification of early warning signals that can highlight suspicious trading at a point in time where the costs of false positives are relatively low (e.g., in the rather lengthy, low-activity accumulation phase preceding a pump). It is important to recognise the presence of both false positives and false negatives in any P&D detection system. In order to minimise the likelihood of Type I errors (i.e., false positives), the parameters for the detection algorithm can be set stricter (e.g., larger price or volume increases) which in turn increases the likelihood of committing a Type II error (i.e., incorrectly missing a real pump-and-dump; false negative). Thus, a cost for both Type I and Type II errors needs to be determined, and a balance struck between the two. The only way to be entirely confident that a particular set of price and volume spikes is the result of a P&D group, is to cross reference those spikes with a group's intent to manipulate. Thus, a desirable area for future research would be to create of a database of confirmed pumps. While labour intensive to do in a fully manual way, the creation of such a database could likely be achieved through a smart combination of automated and manual tasks (e.g., an automated filtering system with human review). Such a database could be used as a means of testing the accuracy

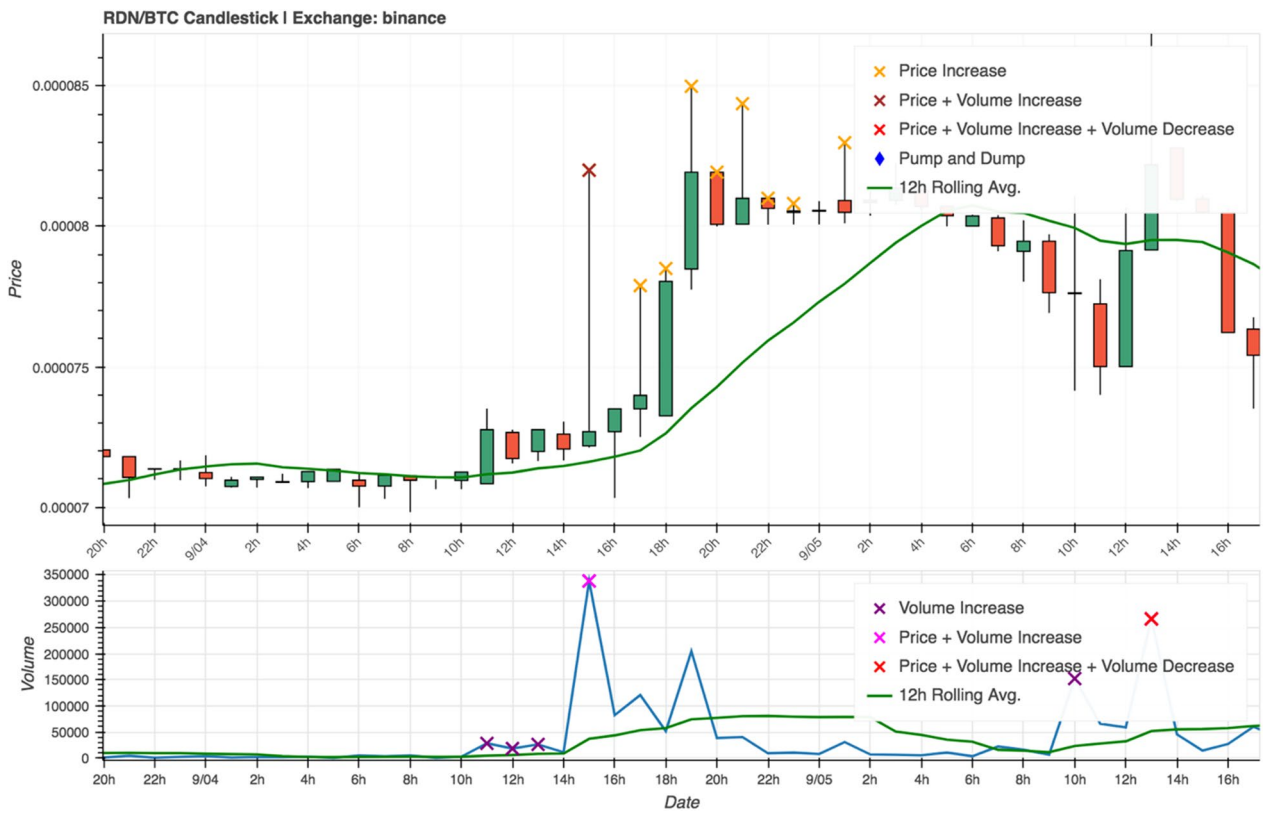


Fig. 11 The chart depicts the results of a pump-and-dump promoted by the group *Moonlight Signal*, which was signalled to commence at 3:30 p.m (UTC) on the 4th of September. While our system correctly marked the corresponding price and volume spikes at the specified time, it failed to identify them as being the result of a pump-and-dump. *Symbol:* RDN/BTC. *Exchange:* Binance

of a detection algorithm, as well as allowing for the use of supervised machine learning methods.

Future research

Two lines of research seem particularly interesting for an extension of cryptocurrency pump-and-dump identification. First, identifying vulnerable coins and understanding the characteristics of those coins that are repeatedly targeted in more detail would allow for efficient resource allocation of detection systems (e.g., those involving both automated systems and human judgment). Second, moving away from exchange trading data, the modus operandi of pump-and-dumps could be examined in more detail. A particularly promising path for future studies could be the linguistic analysis of the coordination of pump-and-dumps in online chat groups, on the one hand; and the means by which misinformation about specific coins is spread on, for example, social media, on the other hand.

Conclusion

This paper has attempted to provide a first look into research for cryptocurrency pump-and-dump schemes. A historical basis for the phenomenon was described with literature from traditional economics and synthesised with the currently available information on cryptocurrency P&D schemes. We proposed a set of defining criteria that could help describe a crypto P&D and showed how an anomaly detection technique could be used to detect patterns of suspicious activity. Ultimately, it is the hope that the information presented in this paper will serve useful as a basis for further research into the detection of these fraudulent schemes.

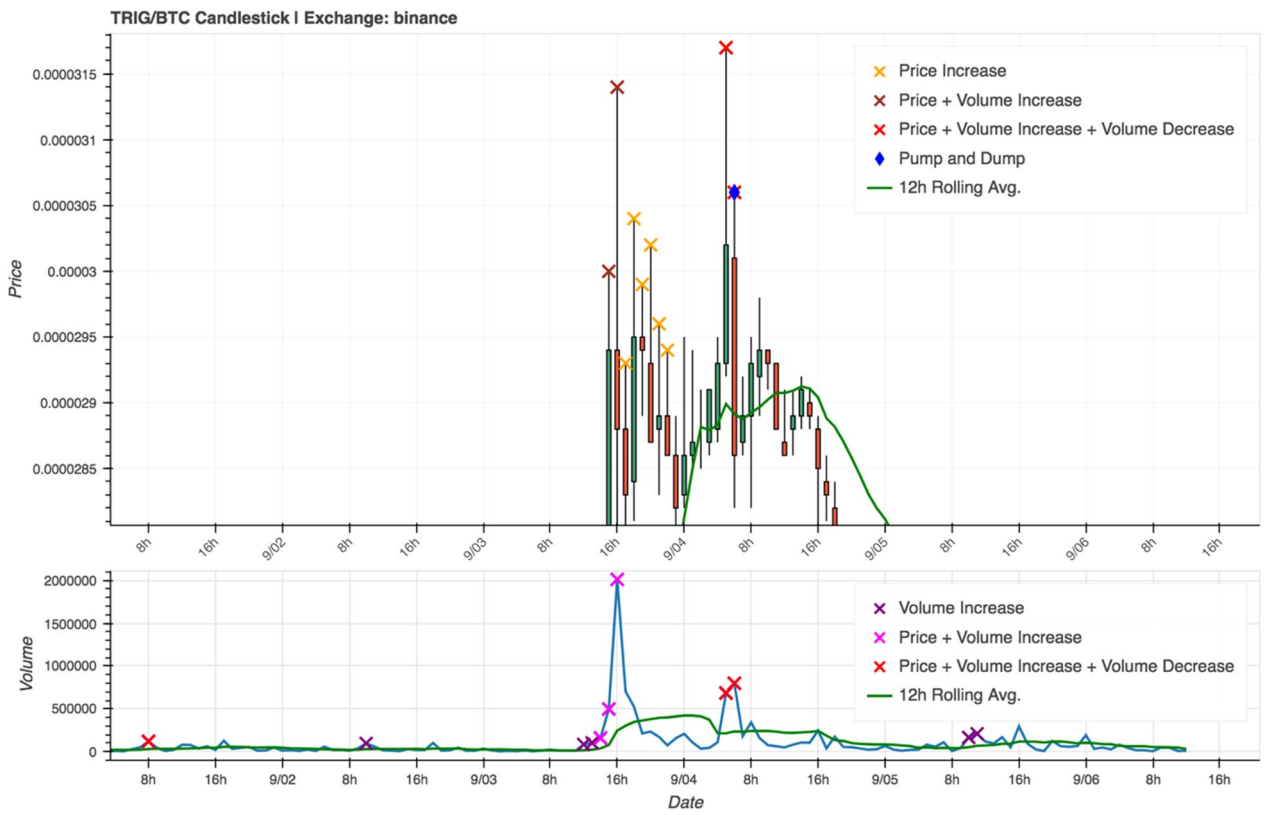


Fig. 12 The chart depicts the results of a pump-and-dump promoted by the group *Crypto Trading*™, which was signalled to commence at 4 p.m (UTC) on the 3rd of September. In this case we once again observe that the system detects large corresponding price and volume spikes at the announced time, however it does not identify these anomalies as being the result a P&D. *Symbol:* TRIG/BTC. *Exchange:* Binance

Abbreviations
Crypto: cryptocurrency; P&D: pump-and-dump; SEC: US Securities and Exchange Commission; OHLCV: Open High Low Close Volume; BTC: Bitcoin; LTC: Litecoin.

Authors' contributions
JK collected the data, ran the analyses and wrote the first draft of the paper; BK and JK conceived of the idea, concept, and analysis; BK wrote the final version of the manuscript. Both authors read and approved the final manuscript.

Author details
[1] Department of Computer Science, VU University Amsterdam, Amsterdam, The Netherlands. [2] Dawes Centre for Future Crime, Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK. [3] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands.

Acknowledgements
Not applicable.

Competing interests
Not applicable.

Availability of data and materials
The data and code needed to reproduce the findings can be found at https://osf.io/827wd/.

Funding
Not applicable.

Appendix

See Table 7.

Table 7 Overview of obtained data between 2018-04-20 00:00:00 to 2018-05-10 23:00:00

Exchange	Symbol pairs	Number of 1 h candles	Days of 1 h history
Binance	302	480	20
Bittrex	242	480	20
Kraken	57	480	20
Kucoin	295	480	20
Lbank	81	480	20

Publisher's Note

Received: 3 July 2018 Accepted: 16 November 2018
Published online: 26 November 2018

References

Bartels, K. C. (2000). Click here to buy the next Microsoft: the penny stock rules, online microcap fraud, and the unwary investor. *Indiana Law Journal, 75,* 353.

Bitcoin Magazine. (2017, May). *What is an Altcoin?* Retrieved from https://bitcoinmagazine.com/guides/what-altcoin/.

Borrion, H. (2013). Quality assurance in crime scripting. *Crime Science, 2*(1), 6. https://doi.org/10.1186/2193-7680-2-6.

Bouraoui, T. (2009). Stock spams: An empirical study on penny stock market. *International Review of Business Research Papers, 5*(4), 292–305.

Brooker, K. (1998, October). *The scary rise of internet stock scams on the net*. Retrieved from http://archive.fortune.com/magazines/fortune/fortune_archive/1998/10/26/250019/index.htm.

Ccxt. (2018). *ccxt/ccxt*. Retrieved from https://github.com/ccxt/ccxt.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR), 41*(3), 15.

Clarke, R. V. (2012). Opportunity makes the thief. Really? And so what? *Crime Science, 1*(1), 3. https://doi.org/10.1186/2193-7680-1-3.

CoinMarketCap. (2018). *Cryptocurrency market capitalizations*. Retrieved from https://coinmarketcap.com/.

Cryptocurrency Prices. (2018). *Compare cryptos to GDP of countries*. Retrieved from http://www.cryptocurrencyprices.net/cryptocurrency_vs_country_gdp.php.

Developments in Banking and Financial Law: 2013. (2014). *Review of banking and financial law*, *33*, 1.

Dugan, B. (2002). The internet and the law part two—Commercial matters: Facilitating and regulating commerce. *Victoria University of Wellington Law Review, 33,* 433.

Farrell, G. (2015). Crime concentration theory. *Crime Prevention and Community Safety, 17*(4), 233–248.

Farrell, G., & Pease, K. (1993). *Once bitten, twice bitten: repeat victimisation and its implications for crime prevention*. London: Home Office Police Research Group.

Keatley, D. (2018). Crime script analysis. *Pathways in crime: An introduction to behaviour sequence analysis* (pp. 125–136). Cham: Springer International Publishing.

Khan, M. F. (2018). *How to avoid getting duped by cryptocurrency pump and dump schemes (like I did)*. Retrieved from https://thenextweb.com/contributors/2018/03/15/avoid-getting-duped-cryptocurrency-pump-dump-schemes-like/.

Kleemans, E. R. (2001). Repeat burglary victimization. Results of empirical research in the Netherlands. In G. Farrell & K. Pease (Eds.), *Repeat Victimization. Crime Prevention Studies* (pp. 53–68). Monsey: Criminal Justice Press.

Kramer, D. B. (2004). The way it is and the way it should be: liability under sec. 10(b) of the exchange act and rule 10b-5 thereunder for making false and misleading statements as part of a scheme to pump and dump a stock. *University of Miami Business Law Review, 13,* 243.

Li, T., Shin, D., & Wang, B. (2018). Cryptocurrency Pump-and-Dump Schemes. Available at SSRN 3267041

Mac, R., & Lytvynenko, J. (2018). *Here's how scammers are using fake news to screw with Bitcoin Investors*. Retrieved from https://www.buzzfeed.com/ryanmac/heres-how-scammers-are-using-fake-news-to-screw-with-bitcoin?utm_term=.ukny8oOev#.iyxxAZ7R3.

Martineau, P. (2018, January). *Inside the group chats where people pump and dump cryptocurrency*. Retrieved from https://theoutline.com/post/3074/inside-the-group-chats-where-people-pump-and-dump-cryptocurrency.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Retrieved from https://bitcoin.org/bitcoin.pdf.

Shifflett, S. (2018, August 05). *Some traders are talking up cryptocurrencies, then dumping them, costing others millions*. Retrieved from https://www.wsj.com/graphics/cryptocurrency-schemes-generate-big-coin/.

Siris, V. A., & Papagalou, F. (2004). Application of anomaly detection algorithms for detecting SYN flooding attacks. In *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE* (vol. 4, pp. 2050–2054). IEEE.

Song, X., Wu, M., Jermaine, C., & Ranka, S. (2007). Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering, 19*(5), 631–645.

Temple, S. (2000). Cybertrading: Financial markets and the internet. *Australian Law Librarian, 8,* 337.

Thompson, P. (2018, June). *Pump and dump in crypto: cases, measures, warnings*. Retrieved from https://cointelegraph.com/news/pump-and-dump-in-crypto-cases-measures-warnings.

Town, S. (2018, February). *How to spot a pump and dump (and avoid it)*. Retrieved from https://cryptobriefing.com/how-to-spot-a-pump-and-dump-avoid/.

U.S. Commodity Futures Trading Commission. (2018). *CFTC issues first pump-and-dump virtual currency customer protection advisory*. Retrieved June 27, 2018, from https://www.cftc.gov/PressRoom/PressReleases/pr7697-18.

US Securities and Exchange Commission. (2017). *Microcap fraud*. Retrieved June 27, 2018, from https://www.sec.gov/spotlight/microcap-fraud.shtml.

Warren, S., Oxburgh, G., Briggs, P., & Wall, D. (2017). How might crime-scripts be used to support the understanding and policing of cloud crime?.

Weisel, D. L. (2005). *Analyzing repeat victimization*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Yang, E., & Worden, J. (2015). The treacherous terrain of penny stocks and how firms are attempting to navigate it.