University at Buffalo The State University of New York

Cryptocurrency Market Manipulation:

Using Unsupervised Learning Techniques to Detect Pump & Dump Activity

Group 25: Devansh Chaudhary, Jae Hyeok Kwak, Robert O’Keeffe

University at Buffalo The State University of New York

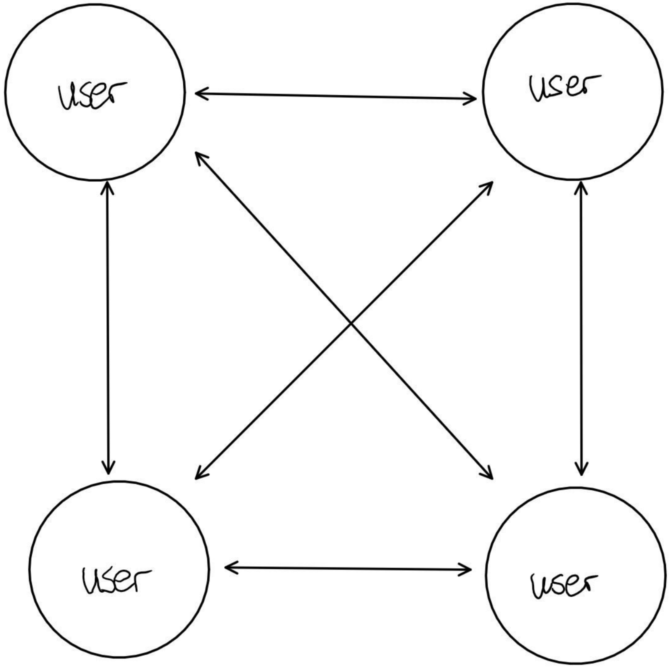
Definitions & Background

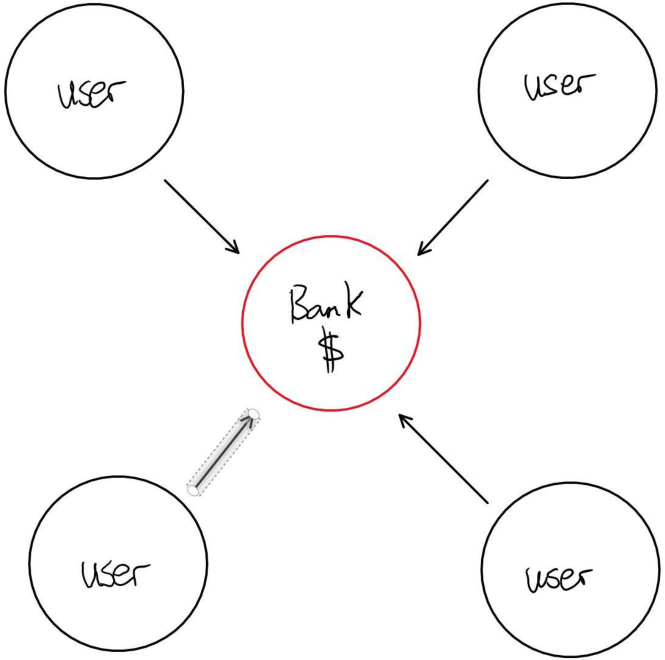
Cryptocurrency

A decentralised Digital Currency which uses the blockchain technology to maintain an online ledger. It uses strong cryptography for secure transactions.

[Click to hear]

BlockChain Technology





[Click to hear]

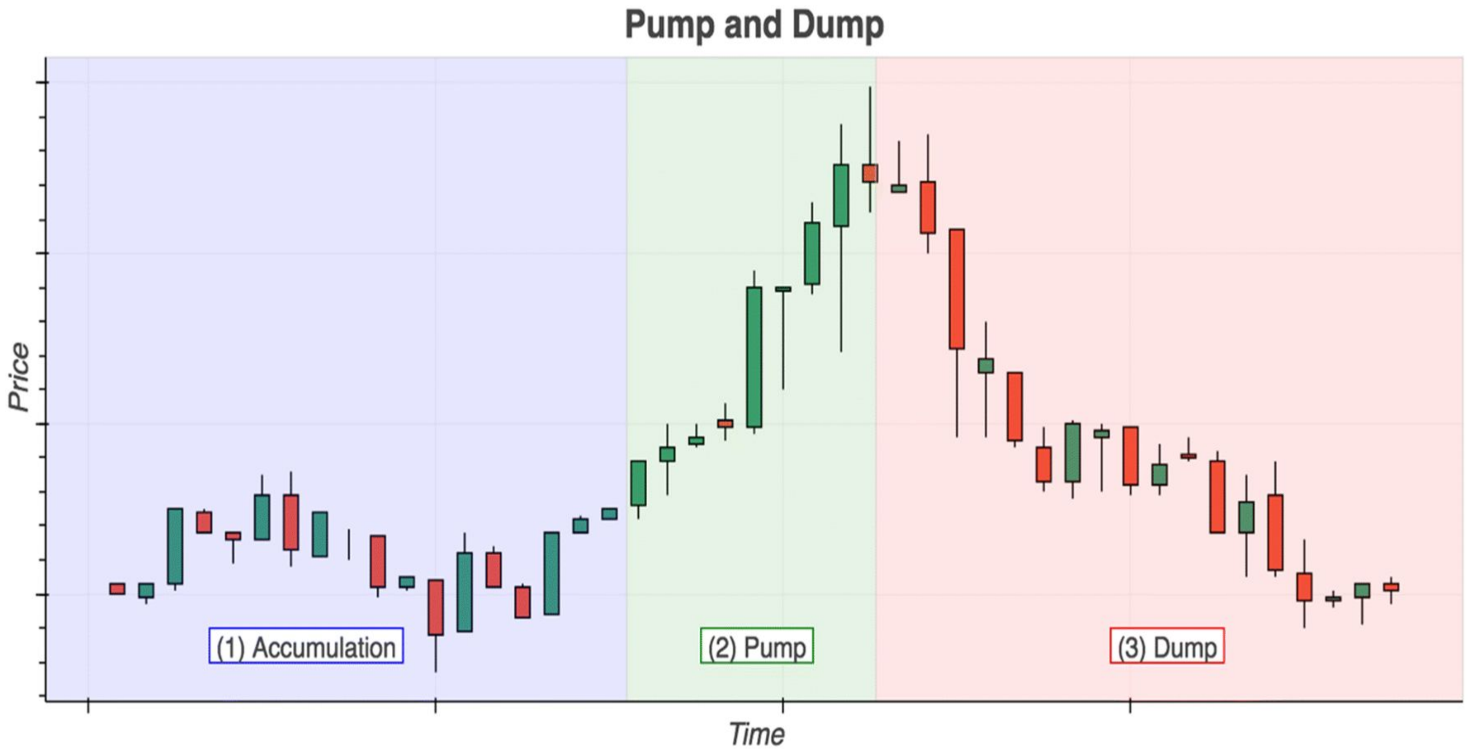
Blockchain transaction

Bank transaction

Crypto Trading

[Click to hear]

Cryptocurrency: Pump and Dump



Click to hear

Unsupervised Methods:

Anomaly Detection & Apriori

2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)

Detecting "Pump & Dump Schemes" on Cryptocurrency Market Using An Improved Apriori Algorithm

Background & Dataset

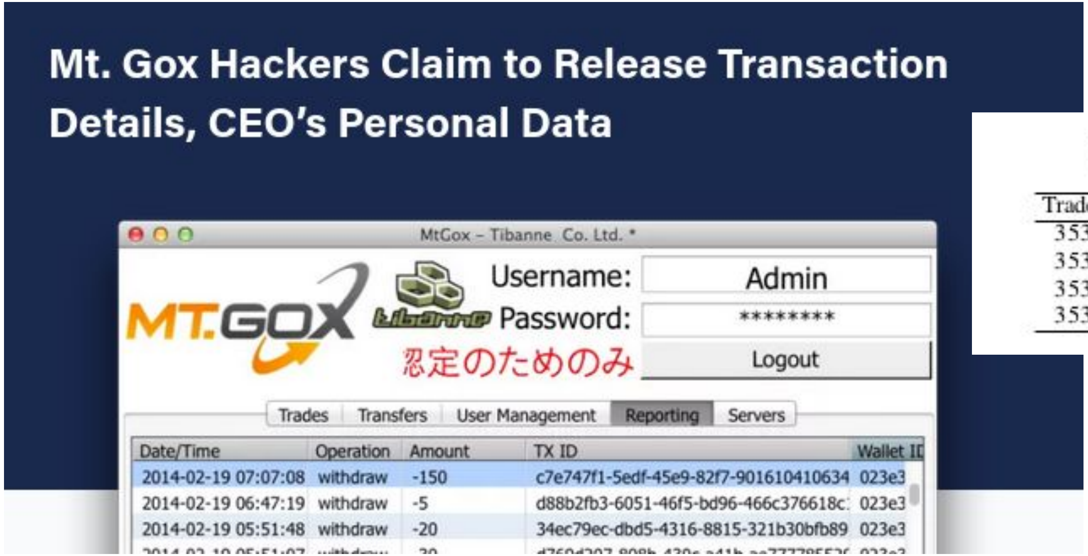


TABLE I: A Segment of The Leaked Data

Trade_Id	Date	User_Id	Type	Bitcoins	Money
35372	2011/04/01 00:28:00	3931	buy	23.02	18.061
35372	2011/04/01 00:28:00	895	sell	23.02	18.061
35373	2011/04/01 00:28:00	722	buy	10	7.8
35373	2011/04/01 00:28:00	895	sell	10	7.8

Apriori Algorithm

	Source	Target	Trade_Id	Bitcoins	Money	Money_Rate	Date
0	895	3931	35372	23.020	18.061	0.784579	2011-04-01 00:28:54
1	895	722	35373	10.000	7.800	0.780000	2011-04-01 00:28:54
2	895	3605	35374	35.000	27.300	0.780000	2011-04-01 00:28:54
3	895	3966	35375	10.600	8.246	0.777925	2011-04-01 00:28:54

	Source	Date_2011-04-01 00:28:54	Date_2011-04-01 06:42:47	Date_2011-04-01 07:21:41
10	2387	0	1	0
11	895	0	0	1
12	895	0	0	1
13	895	0	0	0

Algorithm 1 The improved apriori algorithm

Require: a trading matrix, int mincnt, int span

$C_1 = all\_the\_user$;
$F_1 = \{u \in C_1 | u_{support} \geqslant mincnt\}$;
$supports_1 = \{u_{support} | u \in F_1\}$;
for $k = 2$; $F_{(k-1)} \neq \phi$; $k++$ **do**
  C_k = APRIORI-GEN($F_{(k-1)}$);
  F_k = $\{c \in C_k |$COUNT$(c) \geqslant mincnt\}$;
  $times_k$ = $\{f_{support} | f \in F_k\}$;
end for
return $\{(F_k, times_k) | k = 1, 2, \cdots\}$
function COUNT(c)
  $c = \{user_1, user_2, \ldots, user_n\}$
  T_1={all buy timestamps of $user_1$ };
  T_2={all buy timestamps of $user_2$ };
  ...
  T_n={all buy timestamps of $user_n$ };
  int common_times = 0;
  for all $t_i \in T_1$ **do**
    $t_{i_{begin}} = t_i - span + 1$;
    $t_{i_{end}} = t_i + span$;
    if $i == 1$ **then**
      if $[t_{i_{begin}}, t_{i_{end}}] \cap T_2 \cap \cdots \cap T_n \neq \phi$ **then**
        common_times++;
      end if
    else
      if $t_{i_{begin}} < t_{(i-1)_{end}}$ **then**
        $t_{i_{begin}} = t_{(i-1)_{end}} + 1$;
        if $[t_{i_{begin}}, t_{i_{end}}] \cap T_2 \cap \cdots \cap T_n \neq \phi$ **then**
          common_times++;
        end if
      end if
    end if
  end for
end function

Results

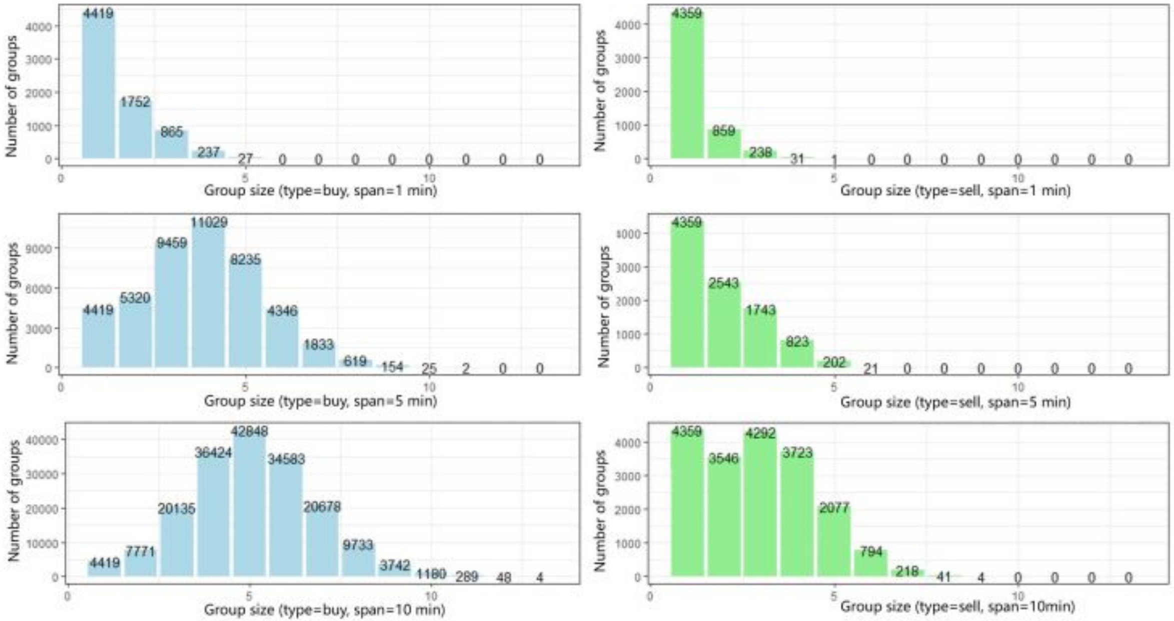


Fig. 2: The Results of Improved Apriori Algorithm.

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

Crime Science

RESEARCH Open Access

CrossMark

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

Background & Dataset

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

Anomaly Detection

$$\mu_y(x) = \frac{\sum_{i=x-y}^{x} x_{close}}{y}$$

$$\mu_y(x) = \frac{\sum_{i=x-y}^{x} x_{volume}}{y}$$

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

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

Results

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% 5%
Parameter price increase	3%	10%	
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

Results

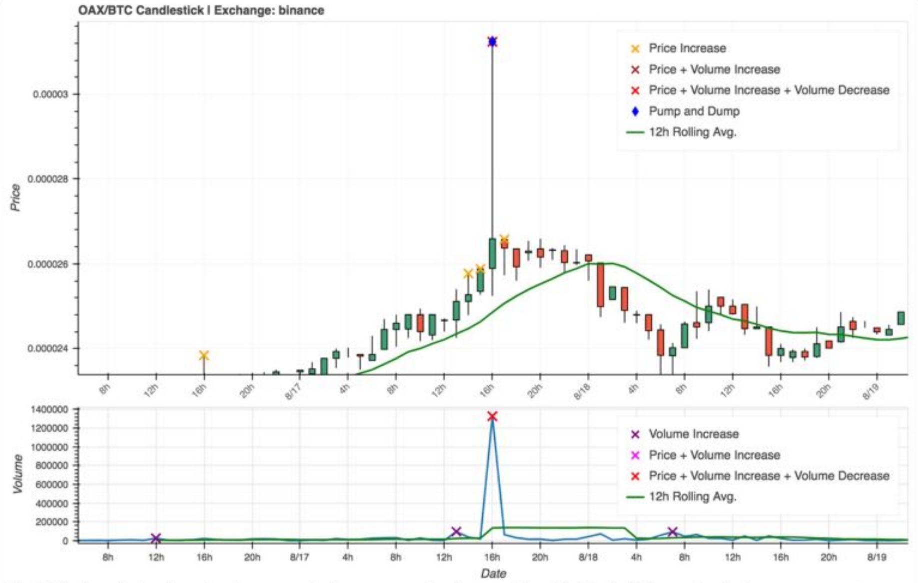


Fig. 9 The chart depicts the results of a pump-and-dump promoted by the group *Moonlight Signal*, which was signalled to commence at 4 pm (UTC) on the 17th of August. Anomalous price and volume spikes at the specified time are clearly visible, and the suspicious activity was correctly marked as a P&D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

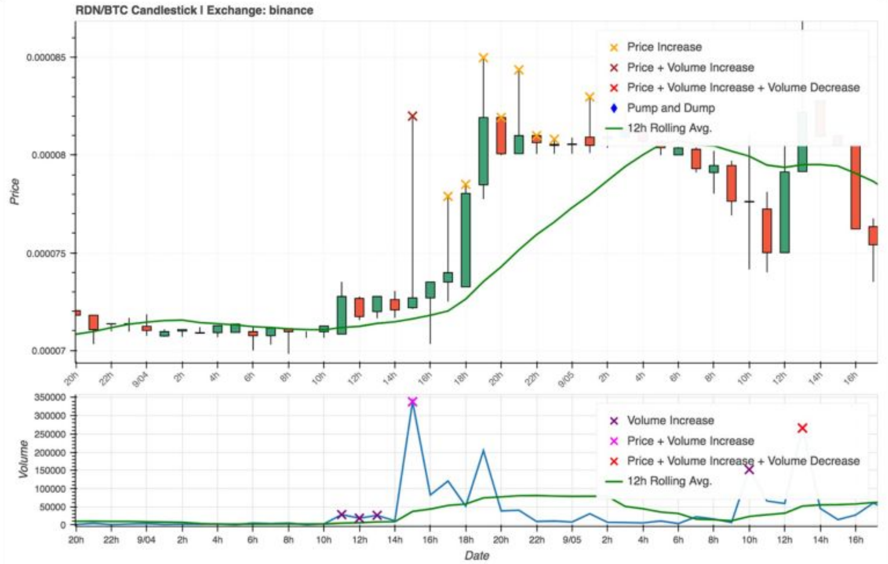


Fig. 11 The chart depicts the results of a pump-and-dump promoted by the group *Moonlight Signal*, which was signalled to commence at 3:30 p.m (UTC) on the 4th of September. While our system correctly marked the corresponding price and volume spikes at the specified time, it failed to identify them as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

Limitations & Looking forward

Limitations: Moving Average

- Based on past data which might not be a good indicator of future changes in demand and supply of the trading asset.

- Moving Averages wont work well during the “Accumulation” phase. Since it takes past price points and returns the average value. The resulting line graph would closely resemble a horizontal line.

- It only works well when the trading asset has a smooth and consistent price range. Since crypto assets are highly volatile. Moving Average might not give accurate results.

Limitations : Apriori

- Computationally expensive as the algorithm has to scan the database repeatedly to find frequent itemsets. It can be taxing if the dataset is large. The dataset used was a portion of all trades. Thus, it did not require much computational power. However, it may not be the case when dealing with all of the data

- Smaller Datasets will lead to the algorithm making false associations mainly because itemset may have higher support and confidence.

Alternative Methods

Frequent Pattern Growth Algorithm

FP Growth Algorithm creates a tree structure to represent the Database. The tree structure is a compressed version of the whole database based on the minimum support count set by the user.

It stores all the transactions and keep track of association between itemsets.

The tree is created by taking all transactions one at a time and mapping them to a path in the tree.

Frequent Pattern Growth Algorithm

TID	*Items bought*	*(ordered) frequent items*
100	{*f, a, c, d, g, i, m, p*}	{*f, c, a, m, p*}
200	{*a, b, c, f, l, m, o*}	{*f, c, a, b, m*}
300	{*b, f, h, j, o, w*}	{*f, b*}
400	{*b, c, k, s, p*}	{*c, b, p*}
500	{*a, f, c, e, l, p, m, n*}	{*f, c, a, m, p*}

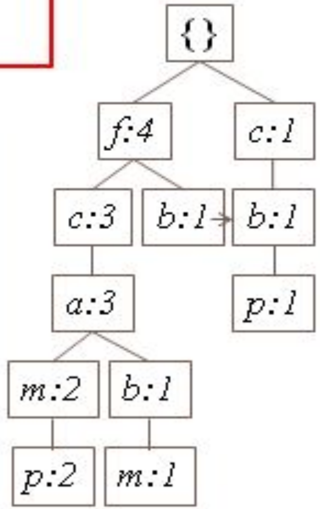
$SuppCount_{Min} = 3$

1. Scan DB once, find frequent 1-itemset (single item pattern)
2. Sort frequent items in frequency descending order, f-list
3. Order items in records
4. Scan DB again, construct FP-tree

F-list=f-c-a-b-m-p

Header Table

Item	*frequency*	*head*
f	4	
c	4	
a	3	
b	3	
m	3	
p	3	



Exponential Moving Average

- It is similar to Simple Moving Average but it adds more weight to more recent data point as opposed to Moving Average which just calculates the Average Price.

- In terms, of trend detection it is has a shorter delay i.e it follows the price more accurately than Moving Average.

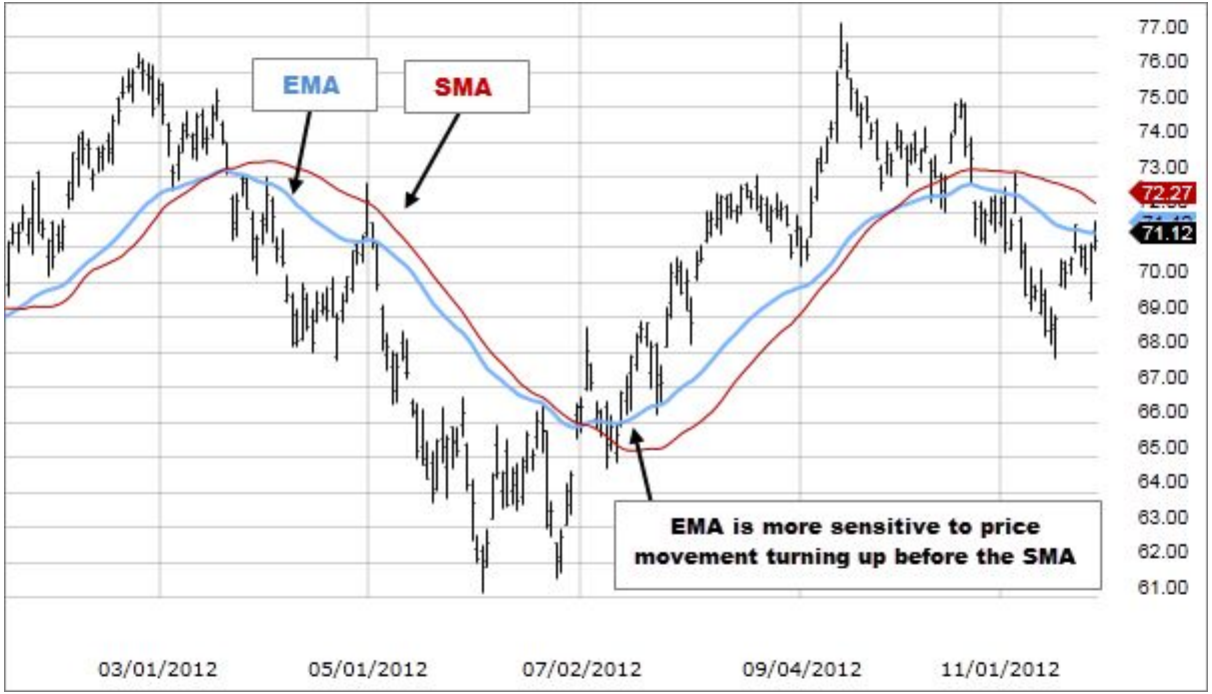
The Formula for EMA Is

$$EMA_{\text{Today}} = \left(\text{Value}_{\text{Today}} * \left(\frac{\text{Smoothing}}{1 + \text{Days}}\right)\right) + EMA_{\text{Yesterday}} * \left(1 - \left(\frac{\text{Smoothing}}{1 + \text{Days}}\right)\right)$$

where:

$EMA = \text{Exponential moving average}$

Exponential Moving Average



Recent Example: Doge Coin April 2021

Dogecoin: From Reddit Meme to Elon Musk's Obsession – the Evolution of DOGE

Dogecoin's explosion in popularity has sent the coin's price soaring 1150% so far this year.

Pump

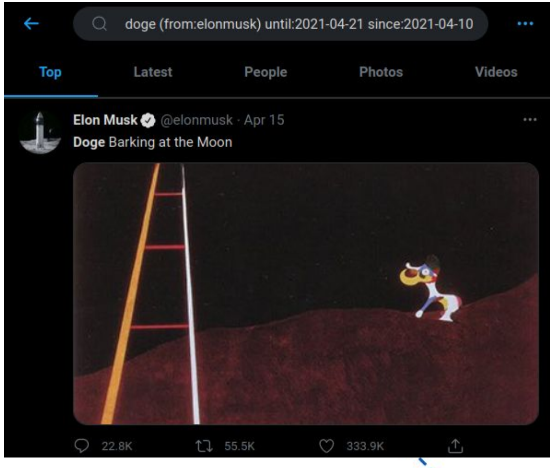
Dump

Dogecoin Disappoints With 32% Dump After 'DOGE Day'

Author: Martin Young • Last Updated Apr 21, 2021 @ 07:24

Crypto meme token Dogecoin has taken a big hit following all the hype leading up to it own celebratory day for fans and holders.


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
Apr 4
2021
Apr 11
Apr 18
Apr 25
May 2
DOGE-USD
2-Day Moving Average
Price Increase Anomaly
Price Decrease Anomaly
Pump Anomaly


doge (from:elonmusk) until:2021-04-21 since:2021-04-10
Top
Latest
People
Photos
Videos
Elon Musk @elonmusk · Apr 15
Doge Barking at the Moon
22.8K
55.5K
333.9K

References

- Kamps, J., Kleinberg, B. To the moon: defining and detecting cryptocurrency pump-and-dumps. *Crime Sci* **7,** 18 (2018). https://doi.org/10.1186/s40163-018-0093-5
- W. Chen, Y. Xu, Z. Zheng, Y. Zhou, J. E. Yang and J. Bian, "Detecting "Pump & Dump Schemes" on Cryptocurrency Market Using An Improved Apriori Algorithm," *2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, 2019, pp. 293-2935, doi: 10.1109/SOSE.2019.00050.

University at Buffalo The State University of New York

End of Presentation

Group 25: Devansh Chaudhary, Jae Hyeok Kwak, Robert O’Keeffe

University at Buffalo The State University of New York