Community impact on a cryptocurrency: Twitter comparison example between Dogecoin and Litecoin

Edouard Lansiaux [1]*,Noé Tchagaspanian[1], Joachim Forget[1,2]

1 Global Variations, 1204 Genève, Switzerland

2 Assemblée nationale, 126 Rue de l'Université, 75355 Paris 07 SP, France

Corresponding author:

*Edouard Lansiaux

E-mail: edouard.lansiaux@globalvariations.com

Twitter: @EdgyLsx

Last name and degree of authors:

Lansiaux, *Msc,* Tchagaspanian, *Msc,* Forget, *MD-PhD.*

Abbreviations:

Contributors: EL and NT designed the research. EL, NT conducted the research. EL did the statistical analysis. EL created the normalized information software. EL and JF wrote the first draft of the paper. EL and JF contributed to the writing of the paper. All authors contributed to the data interpretation, revised each draft for important intellectual content, and read and approved the final manuscript.

Competing interests: All authors declare: no support from any organisation for the submitted work other than that described above; no financial relationships with any organisation that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Word count: 2,400

Number of figures: 4

Number of tables: 0

Supplemental material: 4

Supplementary Figures: 0

Supplementary Tables: 2

Supplementary Files: 2

Keywords: Shannon information entropy; conditional probabilities; Pearson correlation coefficient; Dogecoin; Litecoin; community impact; Twitter

Abstract

Context: The 3rd generation of cryptocurrencies groups together cryptocurrencies as diverse as they are sulphurous, like Dogecoin or Litecoin. While one qualifies as memecoin, the other is of interest to a different category of investors. In our knowledge, no study has independently assessed crypto community economic impact concerning this comparable cryptocurrencies.

Method: Our study has retrospectively studied (from 01/01/2015 to 03/11/2021), using open access data, the association strength (using normalized mutual information) as well as the linear correlation (using Pearson's correlation) between Twitter social networks markers and cryptocurrency economic markers.

Findings and conclusions: While average Dogecoin transaction value is impacted by tweets, Litecoin transactions number and average Litecoin transaction value impacted tweets. Concerning whales, tweets are impacted by Dogecoin whales but any significant relationship was found between Litecoin whales and tweets. Therefore, there are thus the beginnings of a scientific rationale in order to build a trading robot based on these big datas. This paper is only for academic discussion, conclusions need to be confirmed by further research.

Introduction

Since Satoshi Nakamoto's whitepaper in 2008, cryptocurrencies have grown to a huge market capitalisation — currently over $2T. This huge rise in cryptocurrency market capitalisation seems at first glance, deeply linked to the cryptocurrency's community. Indeed, most coins have a strong community promoting it through social networks. One of the most relevant examples when talking about online promotion of a coin might be Elon Musk's tweets. He seemed to have a huge impact on the cryptocurrency market as value seems to increase or decrease as he tweets, which could constitute an insider delay.

But the price of cryptocurrency could in the end be more related to the lindy effect than anything else. According to this effect, the future life expectancy of certain non-perishable goods — such as a technology or an idea — is proportional to their current age. Indeed, the longer something has been around, the more chances there are that it will survive longer. Among memes, competition for survival is fierce. In this jungle the average lifespan is roughly 4 months. When compared to other memes, Doge is kind of a venerated elder. By surviving for eight years — The Doge meme first became popular in 2013 — it has already proven to be one of the most resilient memes of the whole internet history. The Lindy Effect suggests that, for this reason alone, Dogecoin is more likely to persist into the future than any other meme. Just as the US Dollar is backed by America's hegemonic power, Dogecoin is backed by some of the most powerful memes in existence — and the communities behind them. In the meme economy, Doge is as close to a stable thing as you can get. Indeed Dogecoin has a real fan base promoting its use through social networks. Part of what has made Dogecoin a successful cryptocurrency is the non-tribalism of its community.

On the other hand, while being technically very similar (i.e. almost same PoW, and use cases) : Litecoin has a less loud community, despite being older and a more stable cryptocurrency Litecoin doesn't have the same online popularity. Litecoin's users aren't that loud over social networks and do not mean — for most of them — to organise coordinated buying in order to influence the currency's value.

As it was broadly studied and well-documented, crypto economy and wider the financial economy has a big behavioral part with about 20 cognitive biases [1]. Here we will be addressing that question : how much of an impact does online activity — through Twitter in this paper — have on the cryptocurrency market.

Methods

We used two different methods in order to evince association, or lack thereof, between X and Y. These two methods were, on one hand, the classical Pearson correlation and, on the other hand, normalized Shannon mutual information.

Settings

We used historical data, spanning from 01/01/2015 to 03/11/2021, by extracting various economic trackers detailed below .

Variables

For each method, we have studied the following variables:: ‘date (quantitative variable)’, ‘top_100_percent’ 100 first addresses with a large wallet on the studied crypto blockchain (i.e; “whales) (quantitative variable), ‘median_transaction_value’ (quantitative variable), ‘market_cap’ (quantitative variable), ‘average_transaction_value’ (quantitative variable), ‘active_addresses’ on Twitter (i.e. most important influencers) (quantitative variable), ‘tweets’ (quantitative variable).

Data sources

The data frame mainly comes from 3 websites [2-4], but the reason why there are always two data frame versions is that data was lacking for specific days. The first file is the original one which contains some “null” values. But in order to work with our algorithm they have been filled (in the second file) with the average value of the last existing value and next one. This allows us to work with our files without introducing new bias in our correlations.

Statistical method

Obviously, correlation is not causation; but absence of correlation implies absence of causality. Correlation (which might be negative as well as positive) is therefore a key component of the scientific process, for it evinces collections of variables that may interact with each other, hereby warranting further study. Conversely, this methodology also accounts for the early dismissal of unwarranted hypotheses regarding such interplay between variables.

The first method we used is based on the standard Pearson correlation matrix [5], whose computation was performed with the Python Numpy library, that we controlled with two Pearson formulas, for discrete series and continuous series. Specifically, we used the following function:

numpy.corrcoef (df [cols] .values.T)

where:

- df is the dataframe of the data

- cols is the list of columns used for the matrix

About Pearson correlation, it is a commonly formulated criticism that one may not establish a linear correlation between a series of quantitative variables and another one of qualitative variables. However, it will help us to identify those correlations as we are studying them.

The second method we used was based on mutual information entropy [9, 10], allowing us in particular to free ourselves from the limiting assumption of monotony required by linear correlation. It is a measure of the quantity of information (in the sense defined by Claude Shannon in 1948 [11]) that two distributions share. In other words, it is a measure of the association ("clustering") between two variables: it is important to stress the fact that his approach is NOT linear correlation, but classical information entropy. In this approach, we compute a dimensionless quantity generally expressed in units of bits [13], which may be thought of as the reduction in uncertainty about one random variable given knowledge of another. For instance, high mutual information indicates a large reduction in uncertainty about one variable, given the other; whereas low mutual information indicates a small reduction about this uncertainty; and of course zero mutual information between two random variables entails no association between the two distributions. Furthermore, Shannon's source coding theorem establishes strict bounds on what can be known about one data series might be compressed -- which in turn tells how to what extent one variable might be a proxy of another one without data loss. More broadly, Shannon information entropy has been demonstrated to be especially efficacious to evaluate algorithmic complexity when evaluated with the Block Decomposition Method [14,15]. Moreover and according to N. N. Taleb, entropy metrics solve practically all correlation paradoxes in the field of social sciences (or rather, pseudo-paradoxes) [16]. Another important example of the relevance of this technique is that of mother wavelet selection, where it demonstrated superior sensitivity to quantify the changes of signal structure than classical mean-squared error and correlation coefficient [17].

Therefore, in order to compute reproducible results, we use the "muinther" R package available on GitHub which uses these two statistical methods [18].

Biases

The first important bias is the community size. Indeed, that could impact Pearson method, more prone to these issues. However, a larger community will be able to reduce the extreme variations of the variables studied (number of tweets). Therefore, for the two methods, we will not be able to compare the raw data of the samples but only the coefficients (from Pearson or from the normalized information theory) between these two cryptocurrencies.

The second bias is the Pearson method. Indeed, by its definition, Pearson's correlation evaluates the linear relationship between two continuous variables. A relationship is said to be linear when a modification of one of the variables is associated with a proportional modification of the other variable. However, if one moves in a monotonic relationship, the variables tend to change together, but not necessarily at a constant rate. In that case, Spearman's correlation would be better.

Data availability

All data generated or analysed during this study are included in this published article (and its repository [19]).

Results

Pearson correlation analysis

LITECOIN

All studied Pearson correlations for Litecoin (Supplementary File 1 and Figure 1.A.) were significant with a p-value under 0.001 excepting the correlation between the Litecoin market cap and the average litecoin transaction value (p-value= 5.938*10^-3).

Tweets have a small negative impact on Litecoin whales behavior (Pearson coefficient=-0.057). They are positively correlated with median Litecoin transaction value (0.449), average Litecoin transaction value (0.2944), Litecoin market cap (0.469), Litecoin transactions (0.296), and Litecoin active addresses (0.376).

Some results are surprising: Litecoin whales are negatively correlated with Litecoin active addresses (-0.398), with transactions (-0.439), with market cap (-0.466), with average Litecoin transaction value (-0.010) but not with median Litecoin transaction value (0.308).

DOGECOIN

All studied Pearson correlations for Dogecoin (Supplementary File 2. and Figure 1.B.) were significant with a p-value under 0.001.

Tweets are positively correlated to all economic variables: with median Dogecoin transaction value (0.534), with average Dogecoin transaction value (0.543), with Dogecoin market cap (0.549), with Dogecoin whales (0.343), with Dogecoin transactions (0.376), and with Dogecoin active addresses (0.430).

Contrary to Litecoin, Dogecoin whales are positively correlated with Dogecoin active addresses (0.405), with Dogecoin transactions (0.436), with Dogecoin market cap (0.520), with average Dogecoin transaction value (0.452) and with median Dogecoin transaction value (0.476).

Mutual information theory analysis

LITECOIN

“Association” analysis

Community tweets are strongly (with a normalized mutual information coefficient of 0.9 at least) associated (Figure2.A. and Supplementary Table 1) to all Litecoin variables but with fluctuant p-values.

Indeed, only few association p-values are significant: Litecoin average transaction value with tweets (p-value=0.0005), Litecoin average transaction value with whales (0.003), Litecoin active addresses with tweets (0.03), Litecoin transactions with tweets (0), Litecoin transactions with Litecoin active addresses (0.0005), Litecoin transactions with Litecoin average transaction value (0.016).

“Causation” analysis

We are going to explore and to emphasize only significant association causality (previously described).

Concerning tweets association with the economical trackers, the Litecoin transactions number impacted tweets [conditional information entropy of tweets given Litecoin transaction (0.637) is higher than the conditional information entropy of Litecoin transaction given tweets (0.070) as the Supplementary Table 1 shows], and the Litecoin average transaction value impacted tweets too [conditional information entropy of tweets given Litecoin average transaction value (0.637) is higher than the conditional information entropy of Litecoin average transaction value given tweets (0.071)].

Concerning others associations, Litecoin active addresses is impacted by Litecoin transactions [conditional information entropy of Litecoin active addresses given Litecoin transactions (0.071) is higher than the conditional information entropy of Litecoin transactions given Litecoin active addresses (0.037)], and the Litecoin transactions are impacted by Litecoin average transaction value [conditional information entropy of Litecoin transactions given Litecoin average transaction value (0.071) is higher than the conditional information entropy of Litecoin average transaction value given Litecoin transactions (0.070)].

DOGECOIN

"Association" analysis

Community tweets are strongly (with a normalized mutual information coefficient of 0.9 at least) associated (Figure2.B. and Supplementary Table 2) to all Dogecoin variables but with fluctuant p-values.

Significant p-values (under 0.05) only concerned some associations: Dogecoin transactions with median Dogecoin transaction value (0.03), Dogecoin transactions with Dogecoin whales (0.003), average Dogecoin transaction value with Dogecoin market cap (0.011), average Dogecoin transaction value with tweets (0), Dogecoin whales with Dogecoin active addresses ($3.41*10^{-11}$), Dogecoin whales with tweets ($3.22 * 10^{-4}$).

"Causation" analysis

We are going to explore and to emphasize only significant association causality (previously described).

Concerning associations between tweets and Dogecoin economical trackers, average Dogecoin transaction value is impacted by tweets [conditional information entropy of average Dogecoin transaction value given tweets (0.861) is higher than the conditional information entropy of tweets given average Dogecoin transaction value (0.048)] and tweets are impacted by Dogecoin whales [conditional information entropy of tweets given Dogecoin whales (0.861) is higher than the conditional information entropy of Dogecoin whales given tweets (0.124)].

Concerning others associations with whales, Dogecoin active addresses are impacted by Dogecoin whales [conditional information entropy of Dogecoin active addresses given Dogecoin whales (0.120) is higher than the conditional information entropy of Dogecoin whales given Dogecoin active addresses (0.029)]; Dogecoin whales are impacted by Dogecoin transaction [conditional information entropy of Dogecoin whales given Dogecoin transactions (0.124) is higher than the conditional information entropy of Dogecoin transactions given Dogecoin whales(0.049)].

Concerning other associations, Dogecoin transactions impacted the median Dogecoin transaction value [conditional information entropy of median Dogecoin transaction value given Dogecoin transactions (0.078) is higher than the conditional information entropy of Dogecoin transactions given median Dogecoin transaction value (0.049)]; and average Dogecoin transaction value is impacted by Dogecoin market cap [conditional information entropy of average Dogecoin transaction value given Dogecoin market cap (0.047) is higher than the conditional information entropy of Dogecoin market cap given average Dogecoin transaction value (0.001)].

Discussion

While average Dogecoin transaction value is impacted by tweets, Litecoin's transactions number and average Litecoin transaction value impacted tweets. Concerning whales, tweets are impacted by Dogecoin whales but any significant relationship was found between Litecoin whales and tweets. Furthermore, this lack of association was clearly observed using one fundamental approach (mutual information theory), resting on wholly different assumptions and principles: one with classical Pearson correlation, the other with novel (in this field) Shannon mutual information. Furthermore, we observed that these two approaches had contradict a few but mainly rather nicely complemented each other, in that Pearson correlation made it possible to study the sign of the correlation (positive vs. negative) while normalized mutual information made it possible to assess the association strength in a finer way, independent from the assumption of monotony required by linear correlation.

We surmised that the main criticism of our work would most likely be grounded in community size. Indeed, a larger and therefore more active community such as that of a "memecoin" could have a greater impact than a weaker community. Also a qualitative impact of some specific tweets by very well known people could be studied in order to better understand how twitter may impact cryptocurrencies values.

Few studies concerning behavioral impact on a cryptocurrency were made. The first concerns exclusively Bitcoin [20]. The second one is a comparison between Bitcoin or Dogecoin [21]; despite its main limitation that Dogecoin and Bitcoin are from different cryptocurrency generations, this study concludes to an unpredictability of prices according to the tweets of a community. Therefore, we carried out the first study of behavioral impact analysis between comparable cryptocurrencies.

References

1. Douziech T. Les mécaniques du consensus [Internet]. zonebourse. 2021 [cited 2021 Dec 4]. Available from: https://www.zonebourse.com/actualite-bourse/Les-mecaniques-du-consensus--36769932/?countview=0&fbclid=IwAR2Iy174eDAU6Y6mTxO6ZP4agG5QlSfDUOzh0J4CwD77BHiOaYQiuvdsoU4
2. https://bitinfocharts.com ,
3. https://fr.tradingview.com ,
4. https://finance.yahoo.com ,
5. Caut J-L, Pébaÿ P, Lansiaux E, Forget J. COVID_Dept [Internet]. 2021. Available from: https://gitlab.com/covid-data-2/lockdown-and-curfew/-/blob/master/code/COVID_Dept.py
6. Lev J. The Point Biserial Coefficient of Correlation. Ann Math Stat. 1949;20.
7. Tate R. Correlation Between a Discrete and a Continuous Variable. Point-Biserial Correlation. The Annals of Mathematical Statistics. 1954;25.
8. Kornbrot D. Point Biserial Correlation. In: Encyclopedia of Statistics in Behavioral Science [Internet]. American Cancer Society; 2005. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa485
9. Pébaÿ P. Bivariate-cardinalities [Internet]. NexGen Analytics; 2021. Available from: https://gitlab.com/covid-data-2/lockdown-and-curfew/-/blob/master/code/src_bivariate-cardinalities.py
10. Pébaÿ P, Caut J-L, Lansiaux E, Forget J. CF_Dept_entopy [Internet]. 2021. Available from: https://gitlab.com/covid-data-2/lockdown-and-curfew/-/blob/master/code/COVID_Dept_entropy.py
11. Shannon C. A Mathematical Theory of Communication. Bell Labs Technical Journal. 1948;27:379–423.
12. Thompson DC, Pebay PP. Parallel contingency statistics with Titan. 2009; Available from: https://www.osti.gov/biblio/993626
13. McDaid A, Greene D, Hurley N. Normalized Mutual Information to evaluate overlapping community finding algorithms. CoRR. 2011;

14. Zenil H, Soler-Toscano F, Kiani N, Hernandez-Orozco S, Rueda-Toicen A. A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity. Entropy. 2016;20.
15. Zenil H. A Review of Methods for Estimating Algorithmic Complexity: Options, Challenges, and New Directions †. Entropy. 2020;22:612.
16. Taleb NN. Fooled by Correlation: Common Misinterpretations Social 'Science'. 2019 Mar; Available from: https://www.academia.edu/39797871/Fooled_by_Correlation_Common_Misinterpretations_in_Social_Science_
17. Wijaya D, Sarno R, Zulaika E. Information Quality Ratio as a novel metric for mother wavelet selection. Chemometrics and Intelligent Laboratory Systems. 2016;160.
18. Lansiaux E, Caut J-L, Forget J, Pébaÿ P. muinther: relationship study between several variables [Internet]. Available from: https://github.com/edlansiaux/muinther
19. https://github.com/edlansiaux/Behavorial-Cryptos-Study
20. Mai F, Shan Z, Bai Q, Wang X (Shane), Chiang RHL. How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. Journal of Management Information Systems. 2018;35(1):19–52.
21. Tandon C, Revankar S, Palivela H, Parihar SS. How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. International Journal of Information Management Data Insights. 2021;1(2):100035.

Figures

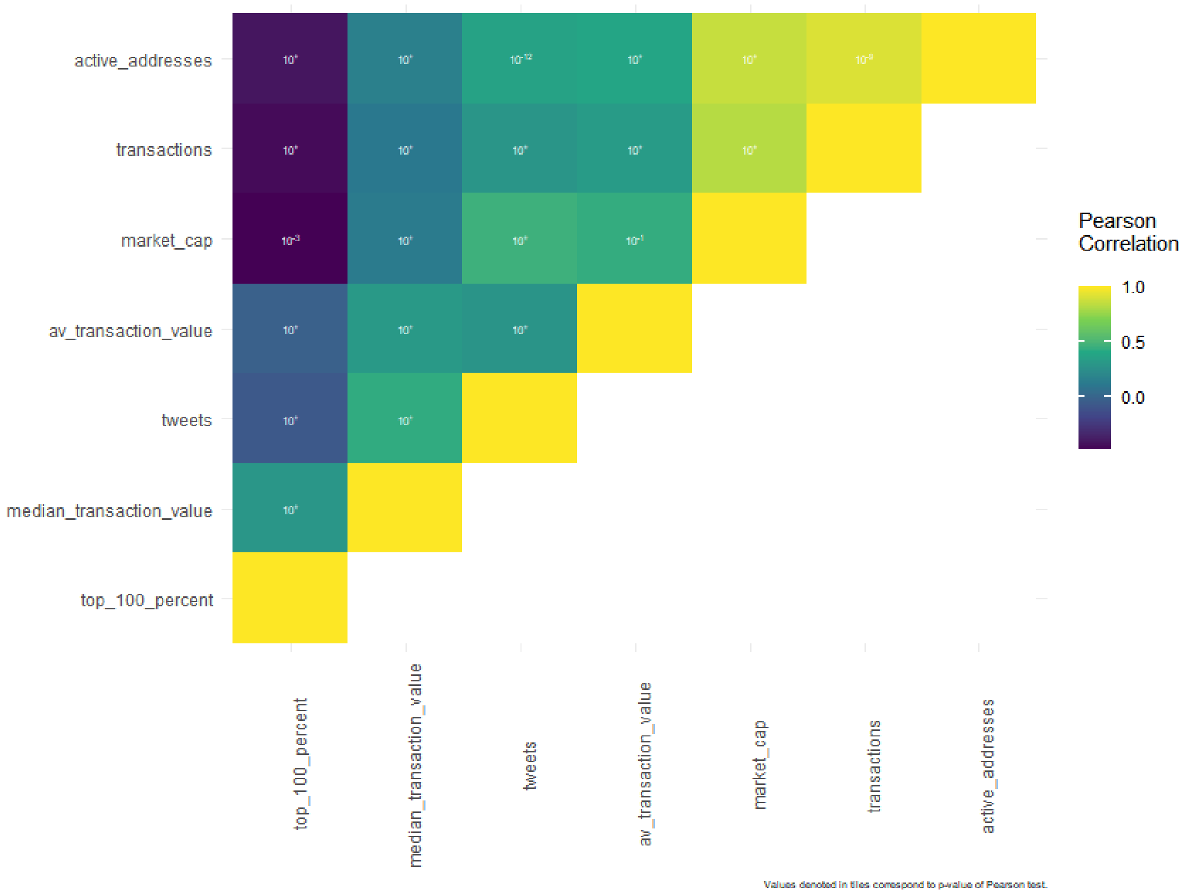

Figure 1.A. Pearson's correlation matrix concerning Dogecoin

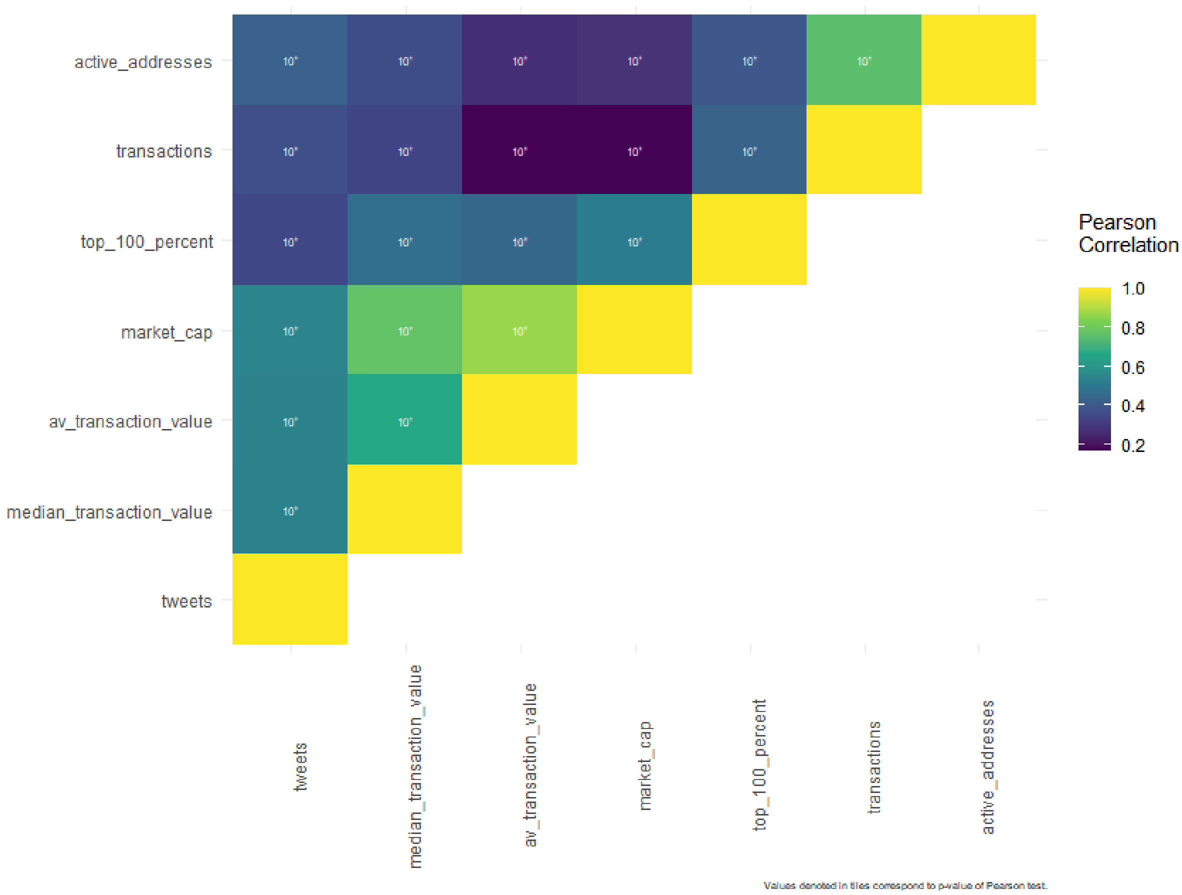

Figure 1.B. Pearson’s correlation matrix concerning Dogecoin

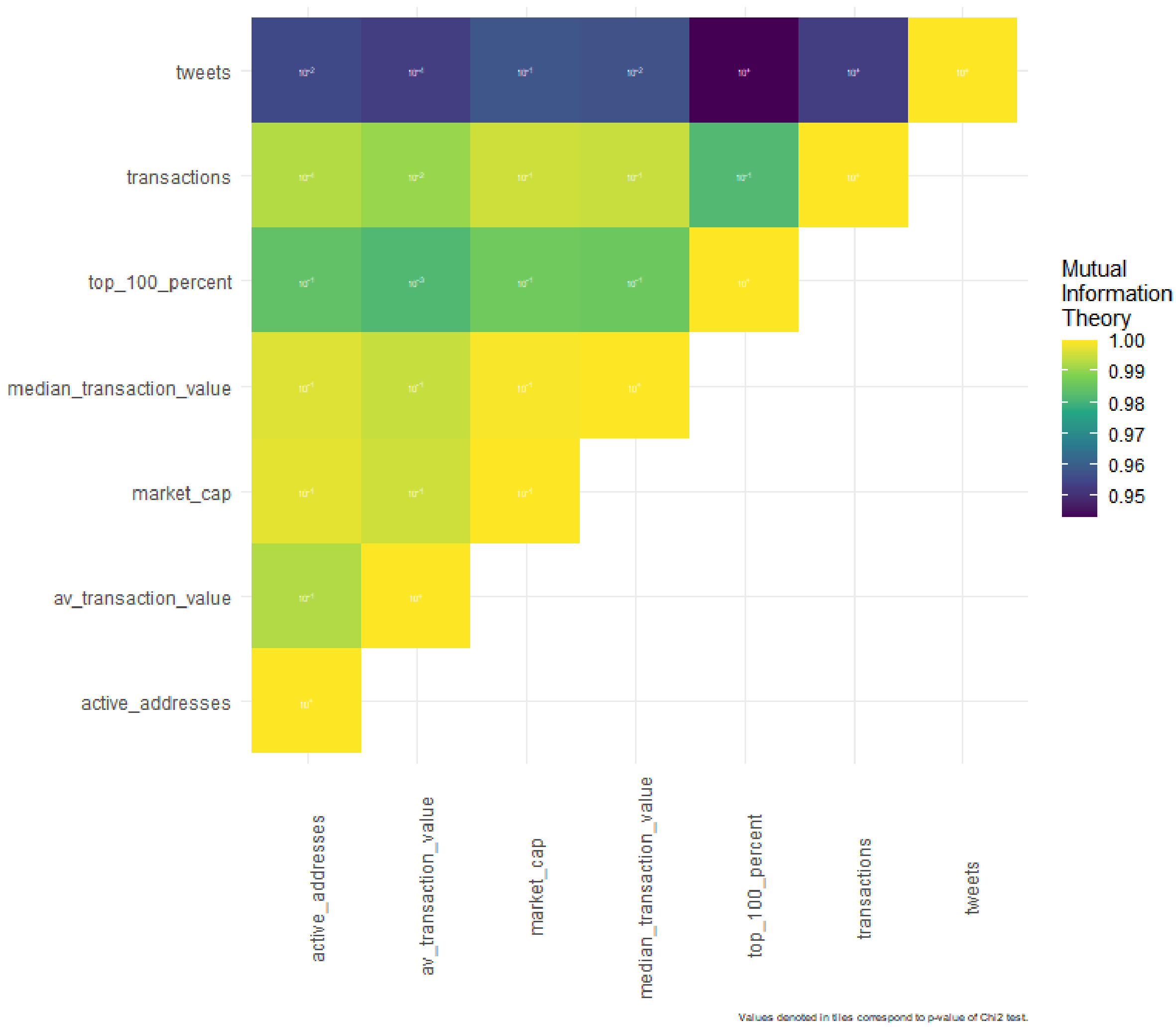

Figure 2.A. Mutual information theory matrix concerning Litecoin

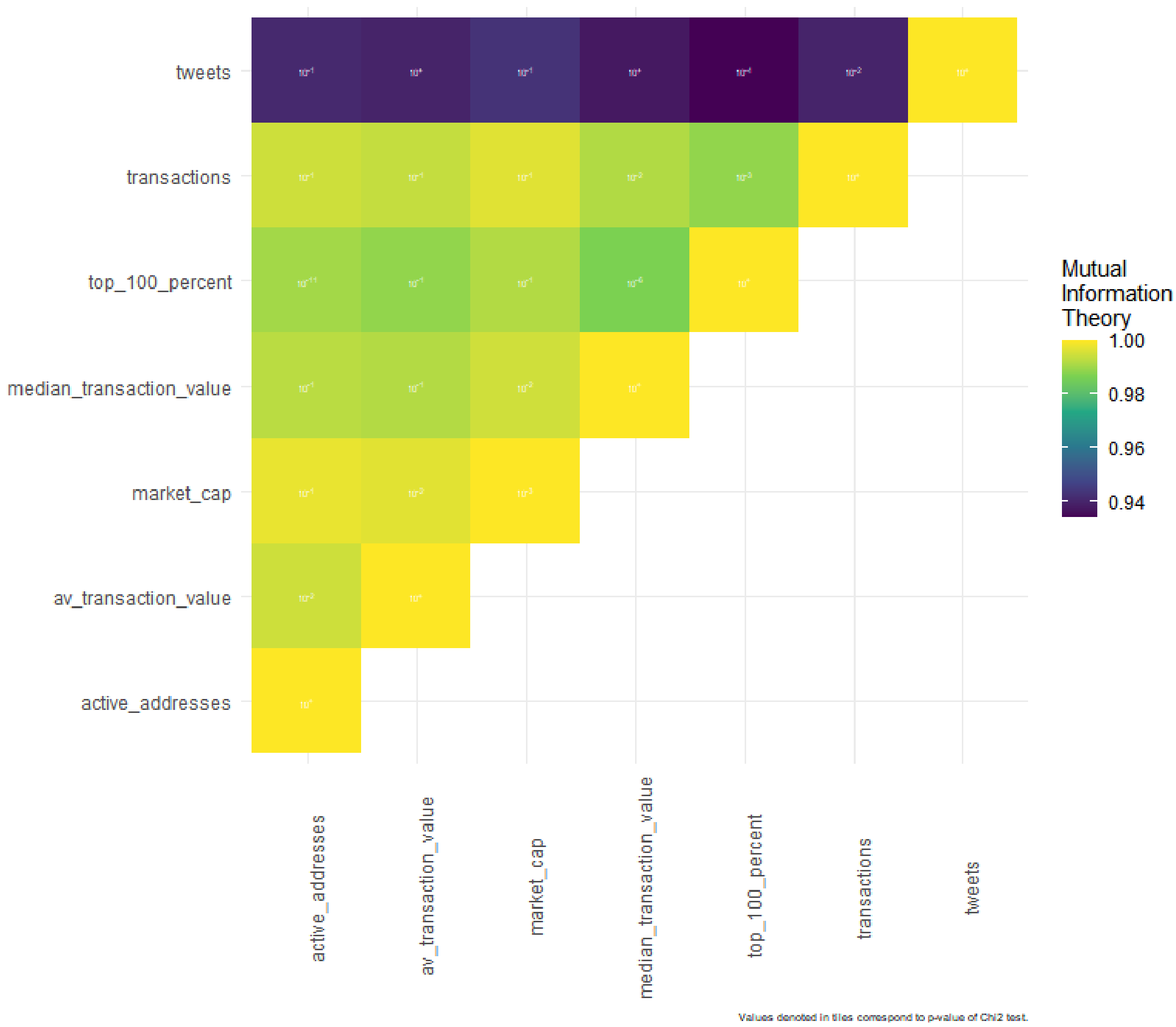

Figure 2.B. Mutual information theory matrix concerning Dogecoin