Is Bitcoin still the representative coin? A network science perspective

Vishwa Shah
DA-IICT
201801036@daiict.ac.in

Pratvi Shah
DA-IICT
201801407@daiict.ac.in

***Abstract*—The aim of this paper is to understand the network structure of the top 100 cryptocurrencies and identify representative coin/s using network parameters. The cryptocurrencies form the nodes and for an edge to exist between two nodes the absolute correlation in that timespan should be greater than the chosen threshold. The community formation and the interactions changing over time bring out the dynamic behavior of the crypto market and allow us to identify how the market changes around the critical events. Further extensive time-series analysis of identified representative coins is performed using Multifractal Detrended Fluctuations Analysis (MFDFA). Unlike the expected largest market cap coin BTC, the majority of network metrics point to ETH as the representative coin and show that the new coins are gaining importance. MFDFA performed on the returns of ETH reflected that they are not independent and possess multifractality due to time-ordering. These results could be useful for investors in understanding the risk-return profile of the cryptocurrency market.**

***Index Terms*—Correlation networks, Cryptocurrencies, Leiden community detection, Overlapping communities, Multifractal detrended fluctuation analysis (MFDFA)**

I. Introduction

Blockchain and Cryptocurrency have piqued the interest of those interested in learning about the financial markets and are intrigued by their decentralized character. These coins provide an alternative asset class for investors wishing to diversify their portfolios due to their high volatility and unique characteristic. However, research in recent years has found a higher correlation between the stock market and the cryptocurrency market [1]. This might be due to cryptocurrency's increasing popularity, which has drawn the attention of long-term investors as well as an increase in investment from traditional financial institutions. This expanding popularity necessitates a greater focus on making the cryptocurrencies' interaction more explainable.

Financial time-series analysis reveals many details ranging from basic statistical analysis [2], [3] to forecasting the price of a given asset [4]. Researchers have utilized publicly accessible transaction records of cryptocurrencies that contain information and traces of financial activities to understand the inherent transaction network structure better [5], [6]. Alongside the network structure, there have been attempts to understand how the crypto market changes around the shocks and to bring out the most influential coins from a network science perspective [7].

Through this paper, we intend to provide insights into the cryptocurrency market from both individual coin perspective and considering it as an asset class. Our paper contributes to the literature in the following ways:

1) Provides a further understanding of the statistical properties of high market cap coins
2) Provides in-depth analysis of the interactions taking place between the coins using dynamic correlation network parameters
3) Brings out the partition and overlapping community structures in which the crypto-market organizes itself
4) Provides insights into the important coins from multiple network parameters perspective
5) Provides extensive analysis of the temporal and spatial fluctuations in important coins' time series around critical events

We start by providing an overview of the dataset in Section II and III, followed by identifying the critical events based on time-series analysis in Section III-B and laying down the complete methodology in Section IV. In Section V, we list down the observations and results obtained from performed experiments. Finally, we conclude the work by summarizing the important results in Section VI.

II. Dataset

We make use of two major types of Time-series datasets:

1) Cryptocurrency price data: This data accounts for the top hundred cryptocurrencies by market cap (from the top 160 cryptocurrencies as we excluded the stable coins and the coins for which no data was available) extracted using polygon API for the timespan of five years (2017- 2022). This dataset consists of opening, closing, highest and lowest price for each day along with the trading volume of that particular coin. BTC and ETH hourly data utilized to generate more finely adjusted observations in Section V-E is fetched from [8].
2) Other asset classes: We include other asset classes like - Oil, Gold, S&P500, and QQQ with cryptocurrency coins to understand their interplay with the cryptocurrency market. We extract the time-series datasets of these asset classes from Yahoo Finance.

III. Exploratory Data Analysis

In this section, we perform basic time-series analysis and also perform the autocorrelation of these series to understand the nature of the data.

A. Basic Statistical Analysis

It has been noted that the way prices fluctuate in the crypto market is similar to how prices change in traditional assets (see Fig. 11 in Appendix A). Furthermore, the returns are more volatile in some time frames, and they have some "calm intervals" (small variations in returns), suggesting the presence of volatility clusters (conditional heteroscedasticity) [9]. This opens up a possibility of returns being correlated to the previous values implying that the returns are not independent of each other.

Table I summarizes the overall descriptive statistics of some of the cryptocurrencies and other asset classes considered in further analysis. These metrics suggest a high deviation from the normal distribution. In general, it is observed that the standard deviation (overall volatility) of the cryptocurrencies is more than that of other asset classes. All other assets, with the exception of Ripple (XRP) and Oil, have negative skewness, which means the left tail is longer than the right tail, suggesting the presence of extreme negative returns. The heavy-tailed distribution is highlighted by the high kurtosis values (normal distributions have kurtosis=3), confirming the existence of outliers in the asset distribution. Furthermore, negative median and positive mean in the case of LTC and XRP suggest the presence of extreme positive returns that led to a mean shift towards the right which is supported by the positive skewness value in the case of XRP.

Upon calculating the metrics on a 90-day sliding window basis we observed that the kurtosis values of coins like ETH, and BTC tend towards normal distribution (towards 3) before 2020 (before the pandemic hit) which might be an indication that these important coins were on their way towards maturity. However, the kurtosis values of almost all assets show a heavy deviation during the pandemic. These values start decreasing after June 2020 suggesting that the market is on its way to recovery.

TABLE I
Summary Statistics for log-returns of high market cap cryptocurrencies and other traditional asset classes.

Asset	Mean	Median	Std	Skewness	Kurtosis
BTC	0.001479	0.001874	0.043106	-0.964271	12.209200
ETH	0.001367	0.001789	0.053922	-0.989688	10.558285
ETC	0.000318	0.000871	0.076911	-0.469832	16.771733
XRP	0.000711	-0.000104	0.067290	0.8789214	13.613933
LTC	0.000521	-0.000092	0.059302	-0.191916	8.1975455
QQQ	0.000567	0.000896	0.011085	-0.327085	10.715326
Oil	0.000724	0.000823	0.025081	1.7936246	42.940031
S&P	0.000371	0.000553	0.009538	-0.331760	22.782824
Gold	0.000217	0.000386	0.007273	-0.631099	9.4202492

B. Identification of Major Events

The high kurtosis value of crypto assets (Table I), suggests the presence of extreme events in their log-returns time-series. We identify these major crashes and peaks to get a closer look at the dynamics of the crypto market around shocks. Since BTC has the highest market capitalization and is assumed to set the trend and govern the market environment, we use its returns time series to identify critical events.

Various research projects have focused on detecting and characterizing such events, although the concept is still up for debate. For event identification, the classification approach described in [12] is employed. These crashes/peaks are sudden, time-localized decrements/increments where normalized returns exceed $\pm 3\sigma'$ or lie near to it, where σ' is the standard deviation of normalized returns.

The returns, log-returns, normalized log-returns and volatility are calculated as given below using Eq. 1-5

$$R_j(i) = \frac{C_j(i)}{C_j(i-1)} \tag{1}$$

$$r_j(i) = \log(R_j(i)) \tag{2}$$

$$r_j(i) = \log C_j(i) - \log C_j(i-1) \tag{3}$$

$$g(t) = \frac{[r(t) - \bar{r}]}{\sigma} \tag{4}$$

$$v_t = \frac{\sum_{t'=t}^{t+n-1} |r(t')|}{n} \tag{5}$$

Here, $R_j(x)$ and $C_j(x)$ represents the return and closing price of j^{th} coin on x^{th} day. $\bar{r}$ and σ represent the mean and the standard deviation of the series $r(t)$ and n is the length of the timespan in consideration (in our case this timespan is 14). A visual representation of the above procedure for BTC time-series along with the explanation of its relation with normalized returns (Eq. 4) and volatility (Eq. 5) is given in the Appendix B for reference.

TABLE II
Common critical events were obtained for BTC, ETH, LTC, ETC, and DOGE. The %Δ is the percentage change in the closing price of BTC immediately before and after the event.

Date	Return	Closing price	%Δ Before	%Δ After
2018-01-16	-0.20	11063.0	-18.59	0.22
2019-09-24	-0.12	8530.01	-12.00	-1.06
2020-03-12	-0.49	4857.1	-38.81	14.88
2020-03-13	0.13	5580.05	14.88	-7.43
2021-05-19	-0.15	36742.3	-14.31	10.56

Critical events in other coins like ETH, LTC, ETC, DOGE, and tokens such as OMG are detected using a similar technique, and the common events identified using all of them are listed in Table II. The dates identified in this sections are then used in later Sections

IV-B, V-B, and V-C to observe the changes in the properties of the network and the multifractal structure of the time series.

C. Autocorrelation

Auto-correlation (AC) is used in addition to the statistical measures listed in Table I, to assess if the trend of returns repeats over time. Fig 1, shows different variations of AC of BTC and S&P time-series where the results of BTC are in accordance with the literature available for financial time-series [10]. There is no significant AC in the returns of BTC and this trend is consistent across other cryptocurrencies. The ACF(Lag) vanishes for almost all lags.

But the presence of volatility clustering suggests that there can be non-linear AC so to test that, AC is performed on absolute log returns. We can observe from the plots in Fig. 1, that there is a significant correlation and in general the ACF(Lag) is slowly decaying. The time series is shuffled for both returns and absolute returns to check if the AC is time-sensitive or is just a random phenomenon. There is no significant AC in shuffled series suggesting time-ordering matters and thus, the returns are not independent across time.

For a long-range correlated series, the ACF(lag) should decline as a power-law function: $ACF(lag) \propto lag^{-\gamma}$ with the exponent $0 < \gamma < 1$. Also, as given in [11], the Generalized Hurst exponent H(2) (a measure of long-term memory of time-series, discussed further in Sections IV-B,V-E) is given as:

$$H(2) = 1 - \frac{\gamma}{2} \tag{6}$$

where γ is the power-law exponent in equation of ACF and lag.

To observe and comment on the presence of these long-range correlations, we perform multifractal analysis on the returns time-series in Section IV-B for some of the representative crypto assets collected from Section IV-A.

IV. Methodology

A. Correlation network Analysis

For our experiments, we have undertaken a similar approach to that mentioned in [13], considering a time period of five years to understand the dynamics of how different coins interact with each other. In order to analyze the correlation networks, we make use of log-returns of each coin as it not only helps us to eliminate the inherent positive bias (trend) in the closing price but also serves as a normalization technique. The missing data is handled by using linear interpolation to estimate the values. We generate dynamically changing networks by dividing the complete time series into smaller timespans and constructing a network for each timespan.

The data for a given timespan is taken to generate the (Pearson) correlation matrix of the coins available. If the absolute value of the Pearson correlation coefficient of returns

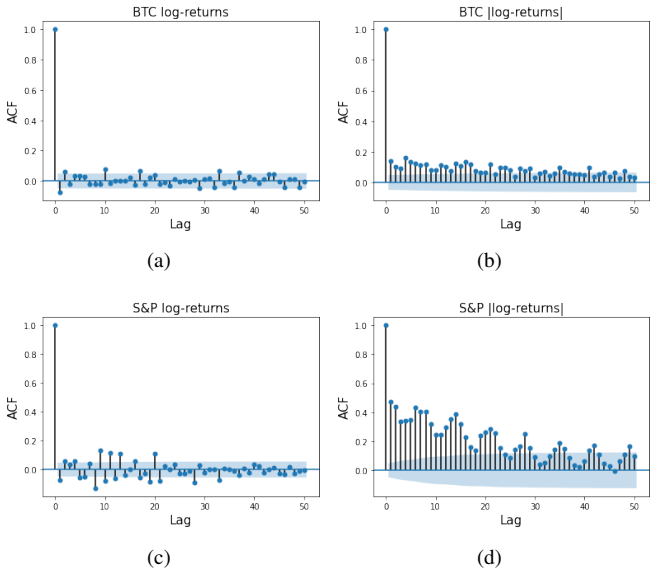


Fig. 1. ACF for different transformations of BTC & S&P returns (a) BTC original returns (b) BTC absolute returns (c) S&P original returns (d) S&P absolute returns

of two coins is greater than the considered threshold then the coins (nodes) are connected in that network. Here, we made an assumption that if there is no data of a coin at the start date of a given timespan then that coin is not included in that network segment. This inclusion of new emerging coins in the network accounts for its dynamic nature.

This dynamic correlation network is then utilized to identify 20% of the coins (17 coins) that have been found up to the year 2018 as in Fig 2. These along with other asset classes QQQ, S&P500, Gold, and Oil forming 21 nodes in total are used to create a static (no addition of new coins) correlation network to analyze the trends and changes in parameters near shocks and major events especially around the time when WHO declared COVID-19 a pandemic (2020-03-11). This will further allow us to understand the interplay between traditional asset classes and the crypto market.

The network parameters of these correlation threshold networks are then analyzed followed by analyses focused on each node. In addition to that, we also look at the inherent community structure of the crypto market and provide an extensive analysis of network properties around the critical events especially COVID-19. This analysis is used to comment on the important coins in the network.

B. MFDFA

From Section III-A, we can see that the time-series of the considered coins do not follow the general normal distribution, and hence, the traditional statistical methods which are true only for fluctuations following normal distribution can not be applied. For complex time series like ours, a more systematic analysis should be employed. Hence, we begin by applying the Detrended Fluctuation Analysis followed by Multifractal Detrended Fluctuation Analysis. DFA is a robust method to

detect long-range correlation even in non-stationary time series and MFDFA is a generalized form of DFA which takes into account complex scaling behavior [14]. For our analysis, we are majorly concerned about identifying the nature of auto-correlation or trend present in the log returns of the representative cryptocurrencies identified in Section IV-A followed by a comment on the presence of multifractal nature in the time series. Monofractal nature implies that the time-series has a constant power-law (scaling) exponent ($h(q)$ =constant in Eq. 7) suggesting the similar effect of different orders of fluctuations but the time-series is said to have multifractal nature when its representation can be explained by scaling exponents $h(q)$ dependent on the order of fluctuation q.

$$F_q(s) \approx s^{h(q)} \tag{7}$$

In Eq. 6 H(2) represents the Hurst exponent h(q) at q=2. If the value of H(2)$>$ 0.5 then the time-series is a long-range dependent (i.e., correlated) structure, and if H(2)$<$ 0.5 the time-series has an anti-correlated structure while equality suggests independent or short-range dependent structure or white noise like variation [15]. Further multifractal characteristics are revealed by the monotonic generalized Hurst exponent $h(q)$, non-linear q-order mass exponent $\tau(q)$, and the singularity spectrum $f(\alpha)$ having the shape of a downward opening parabola. The width of the spectrum $\Delta\alpha$, reveals the degree of multifractality as the higher the width more is the range of exponents in the power-law representation of time series [16]. We can also comment on the efficiency of the market using the MDM value [17]. It uses the deviation of the time series from the random walk as a measure of inefficiency, the higher the deviation more inefficient the market.

TABLE III
STEP WISE EQUATIONS USED IN MFDFA

Function	Value
Original series	x
Mean of original series	μ_x
Detrending	$y_i = \sum_{k=1}^{i}(x_k - \mu_x)$, for i=1,2,... N
Polynomial fitting & RMSE	$\delta y(v,s) = \left\{\frac{1}{s}\sum_{i=1}^{s}[y(v,i) - y_{fit}(v,i)]^2\right\}^{1/2}$
DFA	$F_2(s) = \left\{\frac{1}{N_s}\sum_{v=1}^{N_s}[\delta y(v,s)]^2\right\}^{1/2}$
MFDFA	$F_q(s) = \left\{\frac{1}{N_s}\sum_{v=1}^{N_s}[\delta y(v,s)]^q\right\}^{1/q} \approx s^{h(q)}$
Hurst exponent	$h(q)$
q-order Mass Exponent	$\tau(q) = qh(q) - 1$
Singularity Exponent	$\alpha(q) = \tau'(q)$
Singularity spectrum	$f(\alpha) = q\alpha(q) - \tau(q)$
Width of spectrum	$\Delta\alpha = \alpha_{max} - \alpha_{min}$
MDM	$\frac{1}{2}(\lvert h(-5) - 0.5\rvert + \lvert h(5) - 0.5\rvert)$

The equations used for MFDFA are provided in Table III. For illustrative purposes the step-wise process of MFDFA applied on ETH returns time-series is documented and provided for reference in the Appendix C.

V. OBSERVATIONS AND RESULTS

In this section we start by deciding the timespan for which correlation would be calculated along with the threshold, followed by the results obtained using the methodologies discussed in the previous section.

A. Threshold and timespan

In order to find a good estimate of the necessary threshold and timespan, we performed multiple experiments by varying the threshold for correlation coefficient (from 0.1 to 0.5 in steps of 0.1) and varying the timespan for network creation as 7, 14, 21, 30, 45, and 60 days.

A significant jump in the number of networks having multiple connected components is observed when the threshold is increased from 0.3 (7) to 0.4 (60). So, for our analysis, we have considered 0.3 as the optimal threshold value. After generating the dynamic networks with a 0.3 threshold we observed a log-log linear fit between the total number of nodes and the total number of edges and measured the goodness of fit using the R^2 value (= 1 implies best fit). We found that the R^2 score is between 0.925-0.960 when we vary the timespan keeping the threshold as 0.3. When we increase the timespan the R^2 value decreases till 21 days and thereafter it increases again. Alongside this, we need to consider that when the time span is very large there is a possibility of missing out on the shocks that recover quickly and if the timespan is very small then it might discretize the shock giving an impression of a general trend. So, taking the tradeoff into account we consider a timespan of two weeks (14 days) for all further experiments and perform sliding and moving window analysis Fig 9.

B. General Trends

From Fig 2, we can observe that till 2018, the network is very dynamic in nature leading to a significant change in the network parameters. Before 2018, the average correlation of coins was less than 0.5, and the variance was larger than that after 2018, indicating a more diversified network. Rising correlation leads the market to become more fragile after 2018, with more clustering and a decrease in the characteristic path length, indicating that network characteristics have altered since then. The increase in mean correlation to above 0.5 also illustrates that cryptocurrencies, like any other traditional asset class, are driven by systematic changes, emphasizing the necessity to select a representative for that asset class, whose price movement would provide us with a reasonable approximation of market change.

From Fig 2, we can see that the network properties (density, transitivity, average clustering coefficient, average correlation) tend to follow a similar trend while characteristic path length follows the opposite trend. As a result, focusing on a single property allows us to make general observations about all of the patterns. Fig. 2, shows that the average correlation and density of the network rise directly after the shocks, showing that the network becomes more synchronous after the shocks.

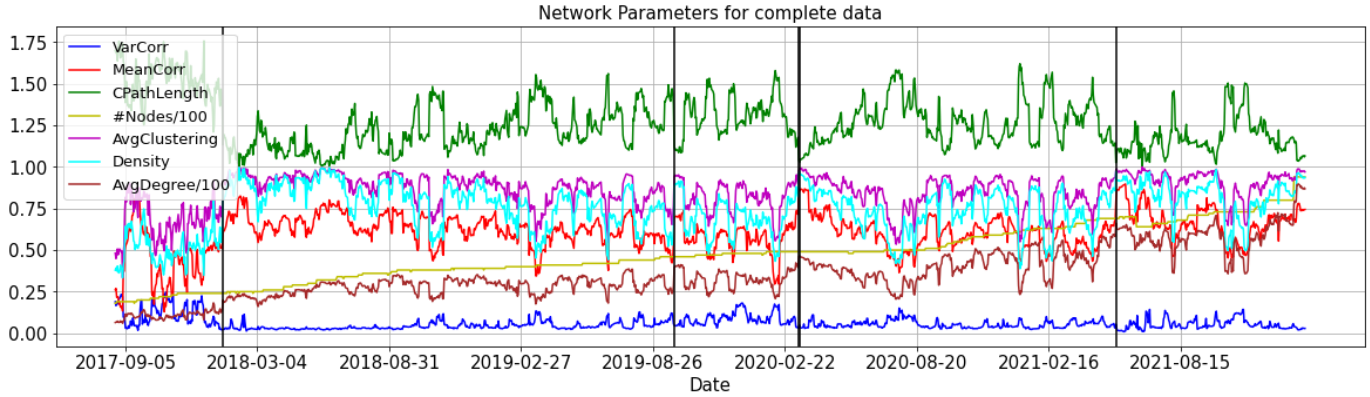


Fig. 2. Network Parameters for complete data with timespan = 14 days and threshold of 0.3. "#Nodes/100" suggests the fraction of coins taken into consideration. "MeanCorr" and "VarCorr" represent the mean and variance, respectively, of the correlation between coins forming the network. Vertical black solid lines represent the critical events given in Table II.

C. Node Metrics

Individual nodes, rather than the entire network, are the subject of this section. To figure out which coins have the most influence on the network in terms of how connected they are to other coins (degree), their local and global influence (eigenvector and Katz centrality), their proximity to other nodes in the network (closeness centrality), and the position in the network (center or periphery), we looked at each network and chose the coins with the highest value of the above parameters. Then we look at how often that coin appears in the complete data. We discovered that BTC, ETH, and XRP were in the top 10 most often occurring coins in both static and dynamic networks, according to all the categories. OMG and BCH are other coins that occurred frequently in the dynamic network while for the static network the other coins are DASH, EOS, ZEC, BNT, XLM, and LTC.

Assessing these parameters individually, it is observed that the centrality measures of BTC, ETH, and XRP remain almost constant when the static network is considered, however, for the dynamic network of all coins it has a decreasing trend suggesting the influence of other newer coins in the network.

D. Community Detection

After observing the network properties and looking at the complete correlation network, we intend to have a closer look at how the coins are organized within the network. The community detection algorithms can be majorly classified into two types: partition (where we get two or more non-intersecting subsets whose union covers all the nodes of the network) and overlapping (two or more intersecting subsets).

For partition-based community detection, the node can belong to at most one community, and hence these techniques can give us information about the different groups formed at a time instant and how these groups (which may have anti-correlated behavior with each other) change with time. For our experiments, we have considered Girwan-Newman [18], Louvain [19], and Leiden [21] community detection algorithms. Due to the relative advantage of higher modularity of Louvain in comparison to Girwan-Newman [20] and connected communities in the case of Leiden over Louvain, we move further with Leiden Communities [21].

TABLE IV
TOP 5 COINS IDENTIFIED USING CENTRALITY VALUES, MAXIMUM OVERLAP, MAXIMUM OCCURRENCE IN THE CENTER OF THE NETWORK, LEAST RMSE FOR THE STATIC NETWORK

Metric	Coins
Degree Centrality	BTC, DASH, ETH, XRP, EOS
Eigen-Vector Centrality	ETH, DASH, BNT, BTC, EOS
Closeness Centrality	DASH, BTC, ETH, EOS, XRP
Center	ETH, LTC, BTC, XRP, EOS
Max Overlap	ETH, BTC, LTC, BNT, OMG
Max Overlap (DegreeC)	ETH, LTC, BTC, DASH, ZEC
Max Overlap (EigenVectorC)	ETH, LTC, BTC, EOS, ZEC
least RMSE	LTC, ETH, ZEC, BTC, DASH

TABLE V
TOP 5 COINS IDENTIFIED USING CENTRALITY VALUES, MAXIMUM OVERLAP, MAXIMUM OCCURRENCE IN THE CENTER OF THE NETWORK, LEAST RMSE FOR THE DYNAMIC NETWORK (100 COINS)

Metric	Coins
Degree Centrality	BNT, DASH, OMG, ETH, XRP
Eigen-Vector Centrality	BNT, OMG, DASH, ETH, QTUM
Closeness Centrality	BNT, DASH, OMG, ETH, XRP
Center	OMG, ETH, DASH, EOS, LTC
Max Overlap	OMG, BNT, NEO, DASH, ZRX
Max Overlap (DegreeC)	ETH, BNT, NEO,TRX, DASH
Max Overlap (EigenVectorC)	ETH, NEO, BNT, EOS, DASH
least RMSE	ADA, QTUM, ETH, TRX, LTC

For this community detection algorithm, we varied the threshold of correlation networks and observed that increasing the threshold helped increased the modularity of the communities detected. We then calculate the root mean square error (RMSE) between the average clustering coefficient (ACC) of the networks (considering the clustering coefficient of all coins that lie in the network at that time) and the ACC of coins belonging to the Leiden community of a particular coin. We calculate this RMSE for all coins and identified coins with

the least RMSE. This allows us to identify the coin/s whose community is a good estimate of the overall network.

After looking at the partition community detection we move on to overlapping communities. For this, we apply the core expansion algorithm of the *cdlib* library, which is an extension of the algorithm proposed in [22] and makes use of the neighborhood overlap and sum overlap method for community detection. Further details of the algorithm are provided in the Appendix D. The nodes appearing in multiple communities act as bridges or information propagators (transmitting and absorbing the shocks).

Overlapping coins (nodes appearing in several communities) are then identified for both static and dynamic networks. These coins are then filtered using Degree Centrality and EigenVector Centrality (coins with a centrality value greater than that of core coins) to ensure that the coins are picked for their worth and importance (as bridges), and not merely because they don't belong to one community (leaf nodes). Additionally, the total networks for which a currency occurs in the overlapped list are counted, and the coins that appeared in the overlapping community the most times are identified. This creates a new technique to identify coins that are essential in terms of network information flow.

Table IV and V lists the top five coins obtained using the above analysis. The representative coin provides us a good approximation of the market and hence, the coin that occurs frequently in all the above analysis can safely be labeled as the representative coin.

BTC, ETH, LTC, and DASH are among the major coins identified in the static network (17 coins and 4 other asset classes). These are the G1 and G2 coins from Table VI. In the dynamic network, we observe the dominance of DASH (G1), ETH (G2) along with some G3 tokens like OMG, BNT (built on ETH's blockchain), and EOS. QTUM (which combines the security of Bitcoin's blockchain model and the flexibility of Ethereum's smart contracts) is also identified among the major coins and is seen as a rising coin. Furthermore, our results are consistent with the results obtained in [7].

E. MFDFA

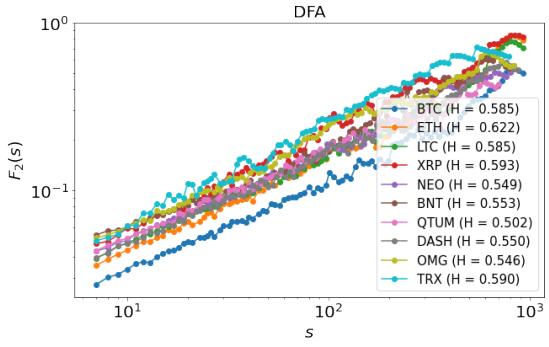


Fig. 3. DFA for multiple cryptocurrencies. Here, q=2 and for this particular case $h(q) = H$ and for any H in [0.5,1] the series is said to be long-range dependent or correlated.

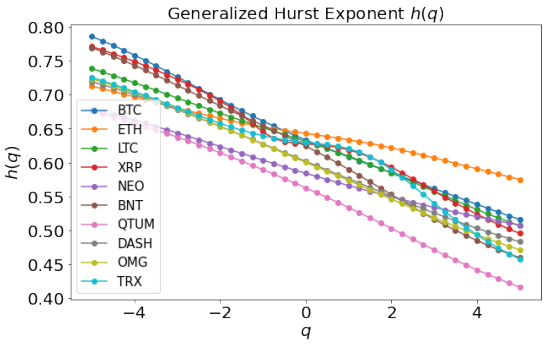


Fig. 4. Variation of generalized Hurst exponent w.r.t the q^{th} order fluctuation. Here, we can see that the Hurst exponent varies with q which suggests that the data is multifractal as for monofractal data the h(q) is constant.

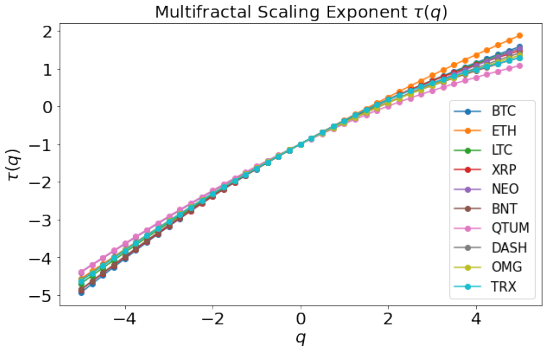


Fig. 5. Multifractal scaling exponent. Here, $\tau(q)$ varies linearly w.r.t. q, when the data is mono-fractal since h(q) will be constant in that case. The non-linear nature of $\tau(q)$ showcases the multifractality of the data.

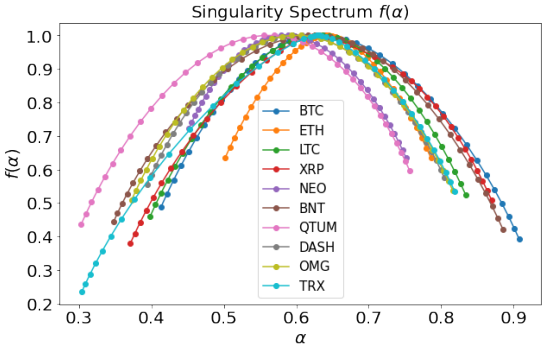
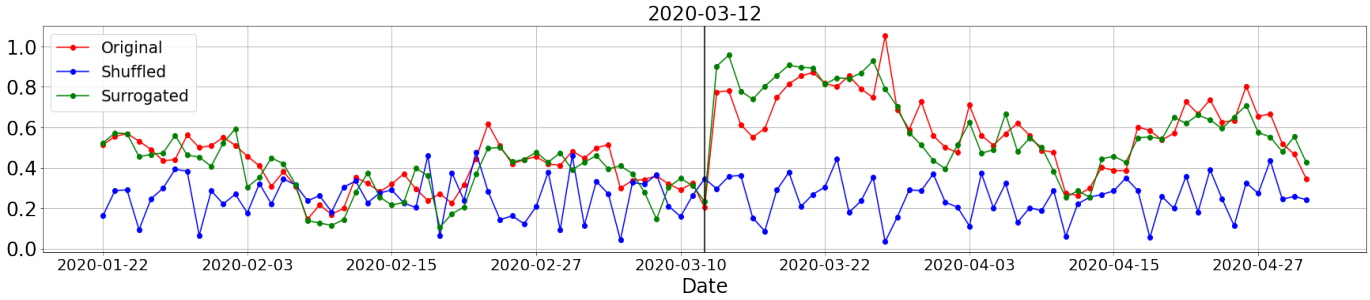


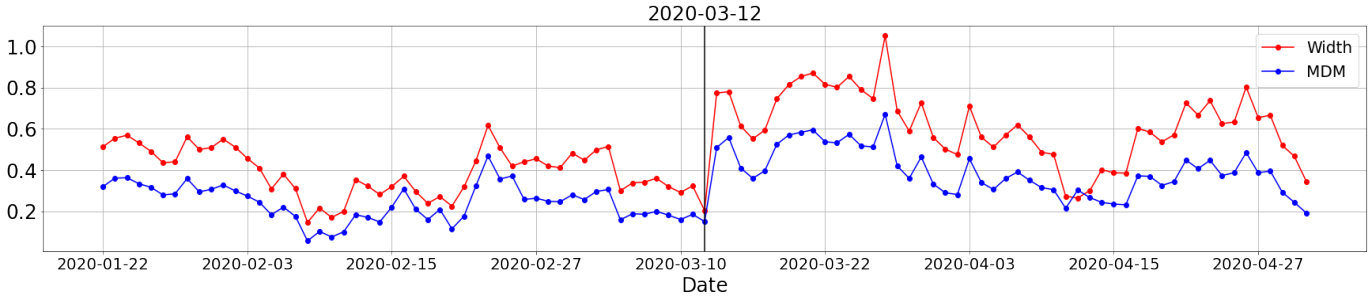
Fig. 6. Singularity Spectrum. The difference between the maximum and minimum α value gives us the multifractal spectrum width $\Delta\alpha$. The larger the width higher is the multifractality of the time series.

In this section, important cryptocurrencies that occur frequently in Table IV and V and the coins which are considered important according to the market capitalization are considered. We have performed MFDFA on the log returns of ETH, LTC, BTC, XRP, NEO, BNT, QTUM, DASH, OMG, TRX.

Fig 3, shows the DFA results with Hurst exponent H for each coin. We can notice that on the log-log scale $F_q(s)$ and s have a linear fit with the slope as H. All the coins have slope(H) > 0.5, and according to equation 6, the power-law



(a)



(b)

Fig. 7. MFDFA performed on hourly data of ETH with a time span of 14 days (336 hours) and sliding it with 24 hours around the critical event of 2020-03-12. (a) Change in width of original, shuffled and surrogated time-series. (b) Change in width and MDM value using original time series.

coefficient γ is slightly less than one, thereby indicating the presence of a weak long-range correlation. This implies that the future trend will be influenced by the series' prior value [23] and the returns time series resembles more to white noise than the random walk [15]. After observing $F_2(q)$, we observe how the generalized Hurst exponent varies with the different order of fluctuations. The non-constant nature of the plot in Fig 4 implies that the time series is sensitive to several sorts of variations (high and low level), demonstrating its multifractal character. This can be further verified by the non-linear nature of Fig 5. From Fig 6, we can see that BNT has the maximum width $\Delta\alpha$ (highest multifractal strength) followed by TRX and XRP whereas ETH has the least multifractal strength.

After confirming the presence of multifractality we check the source of this multifractality. There are two possible sources: (1) long-range correlations of small and large fluctuations within the time series and/or (2) heavy-tailed probability distribution function. Shuffling of time-series removes the correlation or properties due to time ordering and Fourier surrogated time-series destroys the probability distribution [24]. So we created shuffled and surrogated time-series and studied the spectrum to find the source. For shuffled time series, there was a major change in the spectrum, whereas surrogated time series generated a spectrum that was comparable to the original. The width of spectrum produced using hourly data, as shown in Fig. 7(a), reveals this pattern even more clearly. This trend reveals that time ordering is what contributes to the multifractality of the time-series.

So now we take a closer look at how the spectrum and the

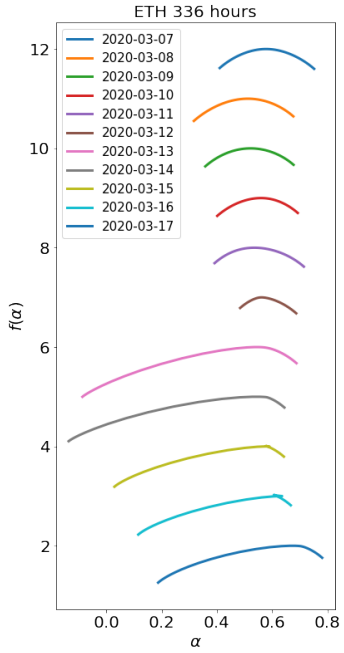


Fig. 8. Change in spectrum around the 2020-03-12 event. The D(α) plots are shifted by one unit for each date for clarity. MFDFA performed on hourly data with $s_{min} = 7$, $s_{max} = 168$ and q in range -5 to 5.

generalized Hurst exponent changes when there is a shock i.e., around the critical events identified in III-B. For this, we resort to the sliding window approach and the focus from hereon will be solely on the coin that was considered to be important from all the network parameters and market capitalization i.e., ETH

and BTC respectively.

Fig. 7(b) and 8 showcases the way width, MDM, and the multifractal spectrum of ETH changes around COVID-19 event. We can observe a sudden increase in the width of the spectrum post the event, accompanied by increase in asymmetry and MDM value (Table III). A similar trend in width and MDM value is also observed around other critical events for both ETH and BTC. This increasing MDM along with increased correlation post-event, as observed in Section V-B, confirms the observation that the market becomes fragile in times of crashes and peaks.

VI. Conclusion

In this work we present a methodology to create a correlation threshold network using the optimal parameters based on the data of top 100 cryptocurrencies and provide multiple techniques to identify the representative or important nodes based on community detection and the network properties. We identify that Ethereum holds a pivotal position in contrast to Bitcoin using the presented methodologies and that new developing coins are considerably influencing network parameters. Further temporal analysis of fluctuations using MFDFA suggested that the time-series of Ethereum and Bitcoin possess a weak long-range correlation with its multifractal nature arising due to time-ordering. Using both network analysis and MFDFA, we can conclude that the efficiency of the market decreases after the critical events with pronounced effect around the COVID-19 crash. So, we can safely say that the event of COVID-19 was the most drastic event in last five years which led to a significant change in not just the equity market but also in the statistical and multifractal properties of the cryptocurrency market.

The network analysis can be further extended from a financial standpoint and the community structure can be utilized to efficiently diversify the portfolio. The work in terms of analysing the nature of asymmetry and the shift in the singularity spectrum (obtained using MFDFA) around critical events can be undertaken as a part of future research.

Acknowledgements

We would like to thank Agam Shah for his consistent support and Dr. Sudheer Chava for allowing us to engage with Georgia Tech's Financial Services Lab. We would also want to thank the Biophysics research group at Lafayette for sharing their thoughts and engaging in useful discussions.

References

[1] https://blogs.imf.org/2022/01/11/crypto-prices-move-more-in-sync-with-stocks-posing-new-risks/

[2] Chan S, Chu J, Nadarajah S, Osterrieder J. A Statistical Analysis of Cryptocurrencies. Journal of Risk and Financial Management. 2017; 10(2):12. https://doi.org/10.3390/jrfm10020012

[3] Vaz de Melo Mendes, Beatriz, and André Fluminense Carneiro. 2020. "A Comprehensive Statistical Analysis of the Six Major Crypto-Currencies from August 2015 through June 2020" Journal of Risk and Financial Management 13, no. 9: 192. https://doi.org/10.3390/jrfm13090192

[4] S. Roy, S. Nanjiba and A. Chakrabarty, "Bitcoin Price Forecasting Using Time Series Analysis," 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018, pp. 1-5, doi: 10.1109/ICCITECHN.2018.8631923.

[5] Serena, L., Ferretti, S. & D'Angelo, G. Cryptocurrencies activity as a complex network: Analysis of transactions graphs. Peer-to-Peer Netw. Appl. 15, 839–853 (2022). https://doi.org/10.1007/s12083-021-01220-4

[6] Wu, J., Liu, J., Zhao, Y., & Zheng, Z. (2021). Analysis of Cryptocurrency Transactions from a Network Perspective: An Overview. J. Netw. Comput. Appl., 190, 103139.

[7] Hong MY, Yoon JW. The impact of COVID-19 on cryptocurrency markets: A network analysis based on mutual information. PLoS One. 2022;17(2):e0259869. Published 2022 Feb 18. doi:10.1371/journal.pone.0259869

[8] https://www.cryptodatadownload.com/data/bitfinex/

[9] Chinthapalli UR. A Comparative Analysis on Probability of Volatility Clusters on Cryptocurrencies, and FOREX Currencies. Journal of Risk and Financial Management. 2021; 14(7):308. https://doi.org/10.3390/jrfm14070308

[10] Cont, R. (2005). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. Capital Markets: Market Microstructure.

[11] Mali, Pritam.A. & Mukhopadhyay, Amitabha. (2015). Long-range memory and multifractality in gold market. Physica Scripta. 10.1088/0031-8949/90/3/035209.

[12] Bielinskyi, A.O., Serdyuk, O., Semerikov, S.O., & Soloviev, V.N. (2021). Econophysics of cryptocurrency crashes: an overview. SHS Web of Conferences.

[13] Burnie, A. (2018). Exploring the Interconnectedness of Cryptocurrencies using Correlation Networks. ArXiv, abs/1806.06632.

[14] Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Bunde, A., Havlin, S., & Stanley, H.E. (2002). Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. Physica A-statistical Mechanics and Its Applications, 316, 87-114.

[15] Ihlen E. A. (2012). Introduction to multifractal detrended fluctuation analysis in matlab. Frontiers in physiology, 3, 141. https://doi.org/10.3389/fphys.2012.00141

[16] Gorjão, L.R., Hassan, G., Kurths, J., & Witthaut, D. (2021). MFDFA: Efficient multifractal detrended fluctuation analysis in python. Computer Physics Communications.

[17] Kakinaka, S and K Umeno [2021] Cryptocurrency market efficiency in short-and long-term horizons during COVID-19: An asymmetric multifractal analysis approach. Finance Research Letters, 102319.

[18] Girvan, Michelle & Newman, Mark. (2002). Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. Proc. Natl Acad. Sci. USA 99, 7821-7826. Proceedings of the National Academy of Sciences of the United States of America. 99. 7821-6. 10.1073/pnas.122653799.

[19] V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech. (2008) P10008, p. 12, 2008.

[20] Rahiminejad, S., Maurya, M.R. & Subramaniam, S. Topological and functional comparison of community detection algorithms in biological networks. BMC Bioinformatics 20, 212 (2019). https://doi.org/10.1186/s12859-019-2746-0

[21] Traag, V.A., Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9, 5233 (2019). https://doi.org/10.1038/s41598-019-41695-z

[22] Choumane, Ali and Awada, Ali and Harkous, Ali,"Core expansion: a new community detection algorithm based on neighborhood overlap", 2020, volume-10, journal of Social Network Analysis and Mining, doi: 10.1007/s13278-020-00647-6

[23] Aslam, F., Mohti, W., & Ferreira, P. (2020). Evidence of Intraday Multifractality in European Stock Markets during the Recent Coronavirus (COVID-19) Outbreak. International Journal of Financial Studies, 8, 1-13.

[24] Kracík, Jiří & Lavicka, Hynek. (2016). Fluctuation analysis of high frequency electric power load in the Czech Republic. Physica A: Statistical Mechanics and its Applications. 462. 10.1016/j.physa.2016.06.073.

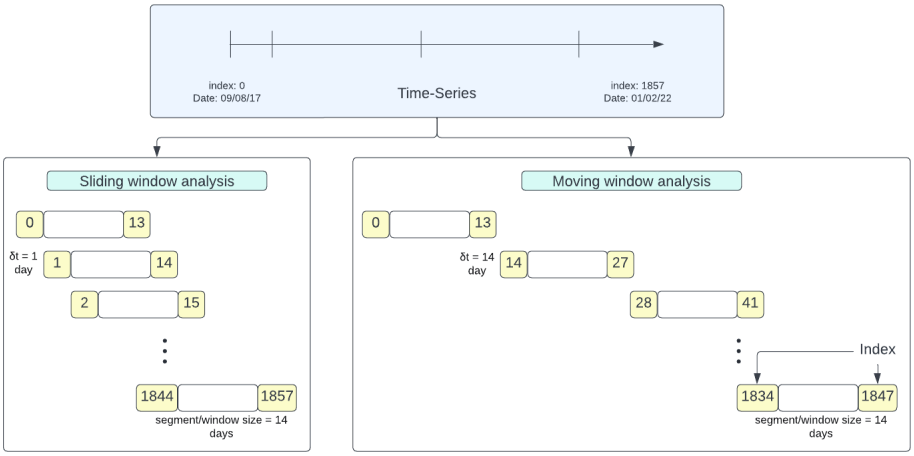
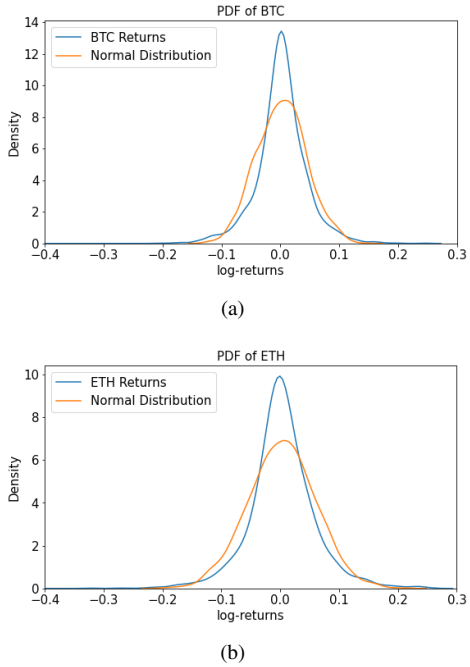


Fig. 9. Sliding (δt = 1 day) and moving time window analysis (δt = 14 days). Here segment/window size is 14 days.

Appendix A
Exploratory Data Analysis

Probability Distribution Plots of Bitcoin and Ethereum are shown in Fig. 10. As seen, both have a fat tail structure. (Kurtosis value $>$ 3). The plots of the closing price and log-returns are shown in Fig. 11

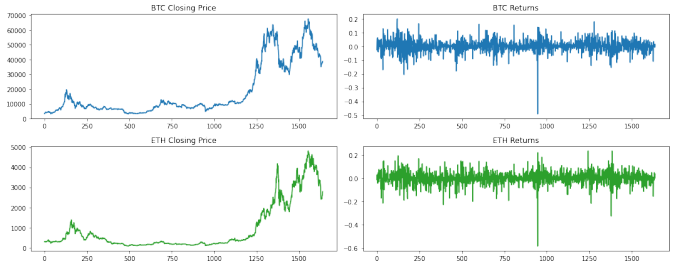


(a)

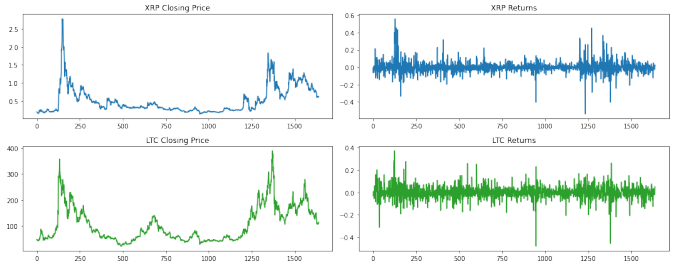
(b)

Fig. 10. Probability distribution plots (PDF) of (a) BTC returns (b) ETH returns. Normal Distribution with the same mean and standard deviation is also plotted to visualize the fat tails represented by the crypto assets

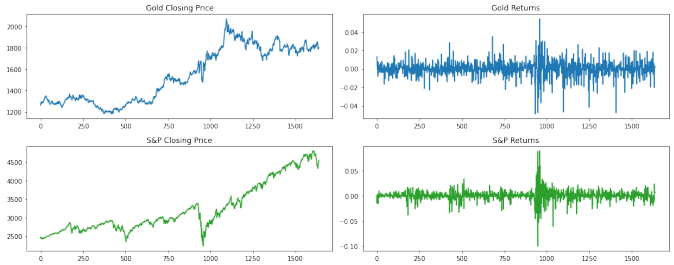
Details about the coins considered in the stable network is given in Table VI.



(a)



(b)

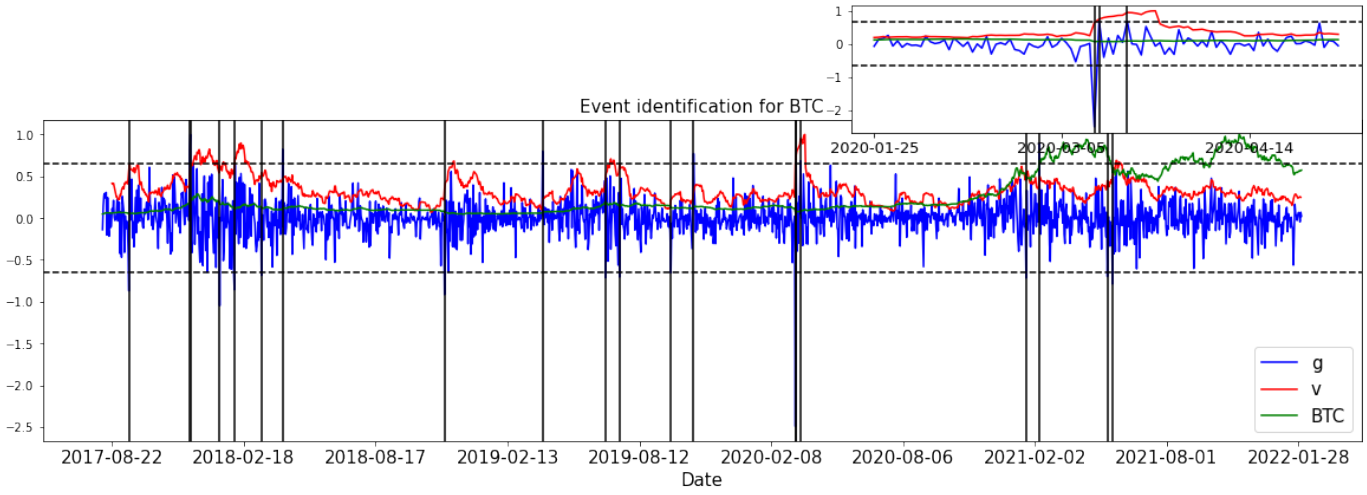


(c)

Fig. 11. Log-returns of (a) BTC & ETH, (b) XRP & LTC, (c) Gold & S&P500.The sudden spike visible in all the above plots is the returns for the day when WHO declared COVID-19 as a pandemic.

TABLE VI
DETAILS OF THE COINS/TOKENS USED TO CREATE THE STATIC NETWORKS OVER TIME

Coin	Chain	PoW vs PoS	Smart Contracts	Coin vs Token	Generation
ETH	ETH	PoW	Yes	Coin	G2
BTC	BTC	PoW	No	Coin	G1
LTC	LTC	PoW	No (Yes since Sep, 2021)	Coin	G1
BNT	ETH	PoA	Yes	Token	G3
OMG	ETH	PoA	Yes	Token	G3
ZEC	ZEC	PoW	No	Coin	G1
DASH	DASH	PoW	No	Coin	G1
EOS	EOS	PoS	Yes	Coin	G3
XRP	XRP	XRP Ledger Consensus Protocol	No (Yes since Aug, 2021)	Coin	G1
XMR	XMR	PoW	No	Coin	G1
ETC	ETC	PoW	Yes	Coin	G2
XLM	XLM	SCP	No	Coin	G1
DOGE	DOGE	PoW	No	Coin	G1
XTZ	XTZ	PoS	Yes	Coin	G3
MANA	ETH	PoW	Yes	Token	G3
XEM	XEM	PoI	No	Coin	G1
STX	BTC	PoX	Yes	Token	G2



The p

Fig. 12. Plot of identified critical events with the variation in normalized returns, volatility and scaled closing price of BTC. The inset shows a closer view of the fluctuation arising around the time when WHO declared COVID-19 a pandemic. As seen the volatility grows on the onset of these events and the normalized returns show deviation from the normal distribution suggesting abnormal behaviour of the market.

APPENDIX B
EVENT IDENTIFICATION

Fig 12 shows the plots of the normalized return of the BTC returns time-series data (blue), calculated volatility (red) as given in Eq. 5 and normalized Bitcoin closing-price series (green). The horizontal dashed lines represent y = $\pm 3\sigma'$ where σ' is the standard deviation of normalized return. The vertical lines are the identified critical events, those events where the normalised returns increases beyond the limits of $\pm 3\sigma'$. We can also observe the presence of high volatility regions around the identified events. In total twenty such events have been identified in the considered time frame (Aug 2017 - Jan 2022).

APPENDIX C
MFDFA

MFDFA method can be divided into the following steps:

1) Detrending: This step is needed to convert the noise like signal/series (x) to a random walk like structure (y). This is incorporated by eliminating the trend (subtract the mean μ_x) and then integrating (cumulative summation in case of discrete time-series).

$$y_i = \sum_{k=1}^{i} (x_k - \mu_x), \text{ for i=1,2,... N}$$

2) RMS variation of timeseries: In this step we have a look at the overall root mean square variation of the timeseries and check whether there are local fluctuations

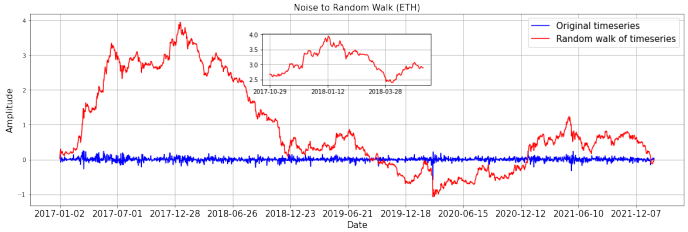


Fig. 13. Conversion of noise like timeseries to random walk like timeseries. The inset shows that when we zoom at the peaks and valleys the nature of timeseries is similar to the complete timeseries indicating the fractal nature of the timeseries.

that have large and small fluctuations or it remains almost constant.

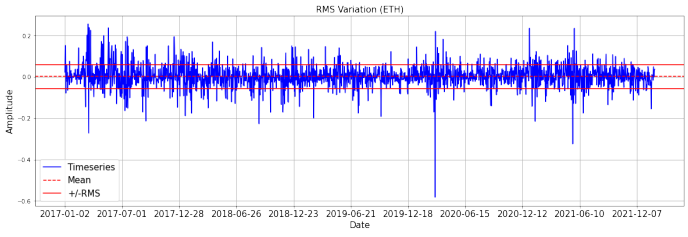


Fig. 14. Overall RMS variation of original timeseries. The mean of the timeseries is zero represented by the dashed red line. The sharp valley near March 2020 and variation within $\pm$ RMS around December 2019 suggests that localized RMS would provide more insights into the nature of fluctuations instead of the overall RMS.

3) Segmenting: In this step, we choose a segment size (s) and divide the complete time-series (N) into N_s=int(N/s) segments to better understand the localized fluctuations.
4) Polynomial fitting and RMSE: Here, we fit an 'm' order polynomial $y_{fit}(v,s)$ to each of the N_s segments and then calculate the RMSE between the original data points and the polynomial fit. This RMSE is referred as $\delta y(v,s)$ where v stands for the v^{th} segment from N_s segments.

$$\delta y(v,s) = \left\{ \frac{1}{s} \sum_{i=1}^{s} [y(v,i) - y_{fit}(v,i)]^2 \right\}^{1/2}$$

5) Detrended Fluctuation Analysis (DFA): From the previous step, we can say that when the segment size is very small then all the fast-moving fluctuations can be captured like in Fig 15(a), 15(b) while the slow moving fluctuations are better captured by the larger segments size Fig 15(c). So, to understand the overall fluctuation we consider the variation (with q = 2) of the RMS over all the segment size.

$$F_2(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [\delta y(v,s)]^2 \right\}^{1/2}$$

6) MFDFA: For understanding the multifractal nature the RMS values from previous step (q=2) are converted to their q^{th} order equivalent by raising them to power of

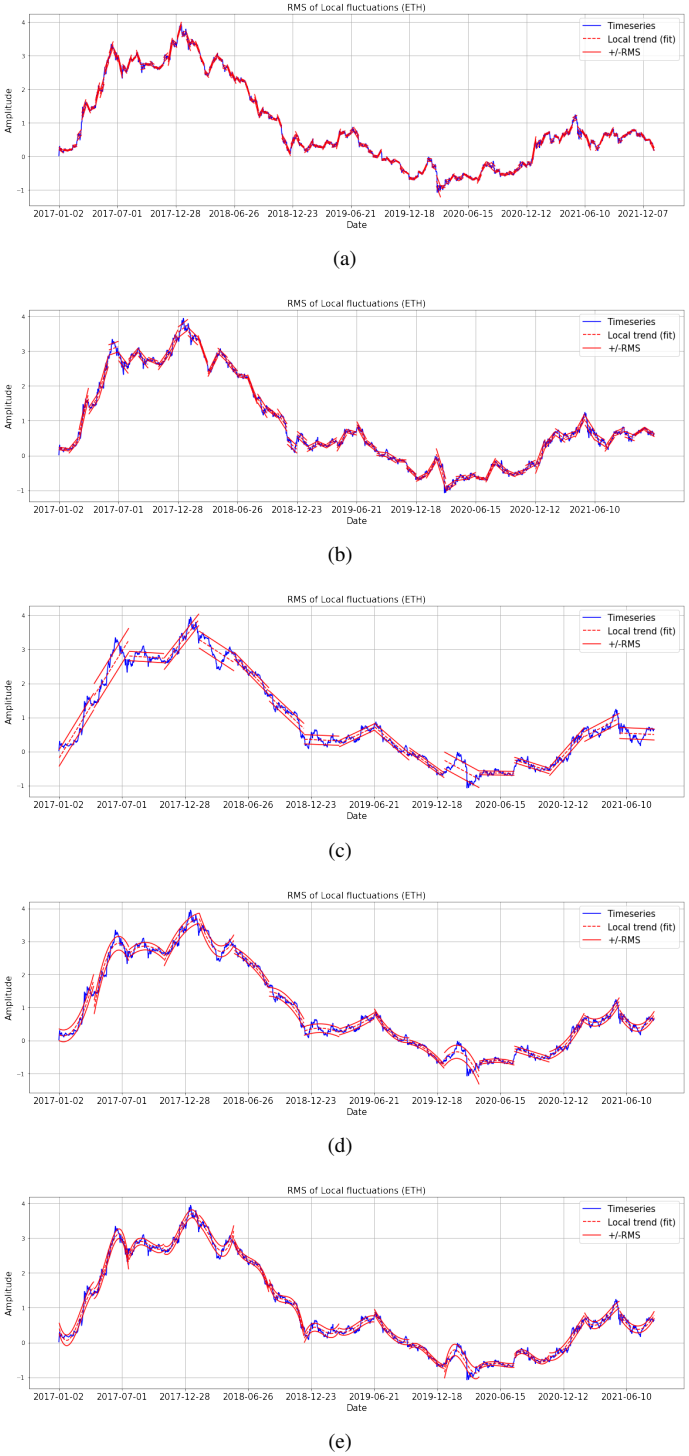


Fig. 15. Polynomial fitting and localized RMS. The timeseries in plots represents the random walk version of the original timeseries (y). Red dotted lines represent the polynomial fit of order m for a duration of s days and the solid red lines around it represents the RMS in that timespan. The order of polynomial fitting can be changed to reduce the RMS and better representation of the fluctuations. For this figures, (a) m = 1, s = 14 days, (b) m = 1, s = 30 days, (c) m = 1, s = 100 days, (d) m = 2, s = 100 days and (e) m = 3, s = 100 days

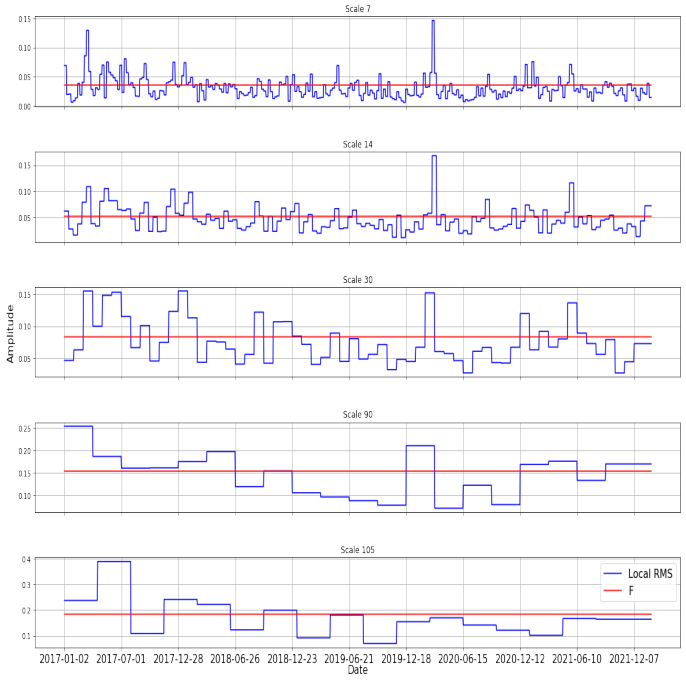


Fig. 16. Variation of second order fluctuation function ($F_2(s)$) with different segment size (s). The y-scale highlights the increase in $F(s)$ value with increasing s.

q. Negative values of q scales the smaller RMS while high positive values of q scales the larger RMS values even more. This makes the overall fluctuation function a better representative of both the smaller and larger fluctuations at the same time. We represent this q^{th} order RMS of $\delta y(v,s)$ over all segments of segment size s as $F_q(s)$.

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [\delta y(v,s)]^q \right\}^{1/q} \approx s^{h(q)}$$

7) Hurst Exponent: The slope of the loglog regression line of $F_2(s)$ v/s s would give us the Hurst exponent (H). In general, the slope of the log-log plot of $F_q(s)$ v/s s would give us the generalized Hurst exponent $h(q)$.

$$F_q(s) \approx s^{h(q)}$$

8) Other parameters for multifractality analysis: In addition to $h(q)$, the mass exponent (multifractal scaling exponent $\tau(q)$) and the singularity spectrum $f(\alpha)$ where the singularity exponent, $\alpha(q) = \tau'(q)$, are also used to determine whether the time-series is multifractal or not. The width $\Delta\alpha$ of the spectrum showcases the strength of multifractality.

$$\tau(q) = qh(q) - 1$$
$$f(\alpha) = q\alpha(q) - \tau(q)$$
$$\Delta\alpha = \alpha_{max} - \alpha_{min}$$

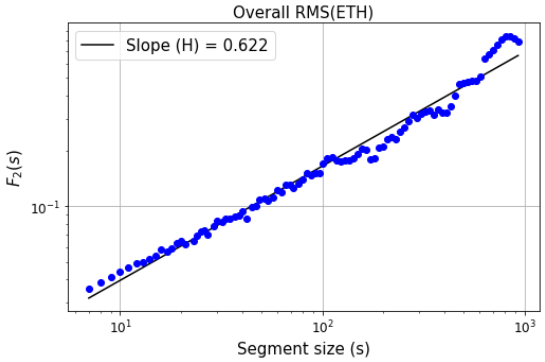


Fig. 17. Overall RMS variation and Hurst Exponent. The blue points represent the value of $F_2(s)$ for a given s and the black line is the regression fit obtained using the log-log variation of $F_2(s)$ and s.

APPENDIX D
CORE-EXPANSION

The core-expansion algorithm works on the principle that nodes belong to the group of other nodes (or a "core") that they share the most neighbors with. The cores are identified in this algorithm using the neighborhood overlap method and the sum overlap method.

Neighborhood overlap is an edge attribute which quantifies the proportion of nodes common between two end points of an edge. Neighborhood overlap between nodes i and j is calculated as:

$$N(i,j) = \frac{c_{ij}}{k_i + k_j - 2 - c_{ij}}$$

Here, c_{ij} is the number of common neighbors of both i and j, and k_i and k_j are their degrees. Higher the ratio (nearer to 1), more are the common nodes between the endpoints i and j of this edge, more the overlap, and more are chances of them being a part of the same community.

After this sum overlap (sum of a node's neighborhood overlaps) is calculated for all nodes and the nodes with local maxima of sum overlap are selected as the core nodes. The core expansion algorithm now works iteratively to assign non-core nodes to their nearest communities based on the greatest overlap. Nodes unassigned to any community, either because they are a terminal leaf node with an overlap value of zero, or because they are evenly split between two communities are then assigned to the community that they have the most edges to. If evenly split, they are assigned to both communities and are the overlapping nodes. These overlapping nodes (coins, in our case), are strongly connected to both the communities (even more than two if these coins have equal edges to more communities detected). Thus, these nodes play an important role in terms of connecting the communities.