

RESEARCH

Analysis of the cryptocurrency market applying different prototype-based clustering techniques

Luiso Lorenzo^{1*†} and Javier Arroyo^{1,2}

*Correspondence: luislore@ucm.es

¹Complutense University, Faculty of Statistical Studies, Madrid, Spain

Full list of author information is available at the end of the article

[†]Equal contributor

Abstract

Since the appearance of Bitcoin, cryptocurrencies have experienced enormous growth not only in terms of capitalization but also in number. As a result, the cryptocurrency market can be an attractive arena for investors as it offers many possibilities, but a difficult one to understand as well. In this work, we aim to summarize and segment the whole cryptocurrency market in 2018 with the help of data analysis tools. We will use three different partitional clustering algorithms each of them using a different representation for cryptocurrencies, namely: yearly mean and standard deviation of the returns, distribution of returns, and time series of returns. Since each representation will provide a different and complementary perspective of the market, we will also explore the combination of the three clustering results to obtain a fine-grained analysis of the main trends of the market. Finally, we will analyse the association of the clustering results with other descriptive features of the cryptocurrencies, including the age, technological attributes, and financial ratios derived from them. This will help to enhance the profiling of the clusters with additional insights. As a result, this work offers a description of the market and a methodology that can be reproduced by investors that want to understand the main trends on the market and that look for cryptocurrencies with different financial performance.

Second part title: Text for this section.

Keywords: Fintech; Data Sciences; Cryptocurrency; Electronic market

Content

The cryptocurrency market consists of more than 4,000 cryptocurrencies^[1] with over 800 trades per second and more than 280 exchanges. It has become a huge new market in a very short term, considering that *Bitcoin* (Nakamoto (2009)), first peer-to-peer and decentralised digital currency was created in 2008 and the first bitcoin was mined in 2009. While cryptocurrencies were originally intended to enable anonymous wire transfers and online purchases, they have become a powerful investment tool.

However, this new market is very diverse. Cryptocurrencies with different technologies, purposes, user base coexist and form a highly heterogeneous market that is difficult to understand and to manage for those addressing a good investment allocation.

^[1]Although cryptoasset is a more general term, as explained in Burniske and Tatar (2017), we will use cryptoasset, cryptocurrency and cryptocurrencies terms indistinguishably in this work.

As other assets, the value of cryptocurrencies swing based on news events, but cryptocurrencies have no physical assets or governments to back their value. Moreover, the cryptocurrency market is new, based on a still developing technology, highly speculative and small in comparison to others. As a result, it is highly volatile with big upswings, bubbles, and sudden market downturns.

Being a market so novel, big, diverse and volatile, it needs to be understood. Several categorization efforts have been made so far. For example, the *Cryptocompare* website ^[2] analyzed over 200 cryptoassets according to regulatory aspects, level of decentralization, supply issuance, economic incentive and others. Such taxonomy is useful even if it only covers approximately the 5% of the existing cryptocurrencies at that time. Another example is in [Burniske and Tatar \(2017\)](#) where classify over 200 cryptocurrencies into three classes of assets based on traditional financial markets, namely: capital asset, consumable/transformable assets and store of value asset. However, this classification is highly subjective as many times the cryptocurrencies may be a combination of some of them. Furthermore, these approaches typically cover a small fraction of the cryptocurrencies, which are the most important ones in terms of volume and popularity, and focus on qualitative aspects or aspects that do not change much.

A different approach consists of analysing the financial performance of the cryptocurrencies and describing it from a statistical point of view. In [Chan et al. \(2017\)](#) analyses a few cryptocurrencies (Bitcoin, Dash, Dogecoin, Litecoin, MaidSafeCoin, Monero and Ripple) which exhibit heavy-tailed distributions that fits the generalized hyperbolic distributions. In [Hu et al. \(2019\)](#) analyses the stylized facts and return properties of 222 cryptocurrencies and find a large degree of skewness and volatility in the population of returns. Furthermore, according to [Pele et al. \(2020\)](#) cryptocurrencies can be clearly separated from classical assets, mainly due to their tail behaviour. However, their cluster results also reveal that the behaviour of the cryptocurrencies is diverse.

The same conclusion can be drawn in other clustering analysis using cryptocurrencies. In [Stosic et al. \(2018\)](#) is represented the correlations of 119 cryptocurrencymarket as a complex network and discover distinct community structures in its minimum spanning tree. In [Song et al. \(2019\)](#) analyses 76 cryptocurrencies using the correlation-based clustering and filtering out the linear influences of Bitcoin and Ethereum and detect 6 clusters, but that do not remain stable after the announcement of regulations from various countries. The time dimension plays an important role in the work in [Sigaki et al. \(2019\)](#), that cluster the time series of 437 cryptocurrencies using hierarchical clustering and detect 4 different groups where the behaviour evolves differently in terms of efficiency for the information.

All these approaches reveal that it is possible to establish different groups of cryptocurrencies in terms of their financial performance. And identifying them, it is useful to better understand the cryptocurrency market, but also for building a diversified portfolio.

All of these methods use different representations of the cryptocurrencies: correlations in [Song et al. \(2019\)](#); [Stosic et al. \(2018\)](#), factors extracted from the correlation

^[2]<https://cryptocompare.com>

matrix (Pele et al. (2020)) and time series (Sigaki et al. (2019)). Each representation focuses on different aspects of the cryptocurrency that are meaningful for the purpose of the analysis.

However, it would be possible to combine the clustering results using different representations of the cryptocurrencies where each one take into account different aspects of the cryptocurrencies. In this way, the combination of the clustering results would make possible to characterize each cryptocurrency in several dimensions, one for each cluster strategy. If the clusters for each cluster strategy are meaningful, their combination would offer a more detailed characterization of the market and useful insights for portfolio management.

In this work, we will explore the combination of clustering of cryptocurrencies using R Core Team (2013). We will go beyond a few hundreds cryptocurrencies, as most studies do, and will explore all the cryptocurrencies traded in 2018 (more than 1,700 cryptocurrencies) easily scalable for a growing market. We will describe each cryptocurrency considering the log return transformation of the daily price in 2018 with three different levels of granularity:

- Mean and standard deviation of the daily returns
- Distribution of the daily returns
- Time-series of the daily returns

In the first case, we provide a meaningful summary commonly used to describe financial assets over time as it is the annualized return and volatility, or with the central tendency and the dispersion of the returns. In the second case, we consider the whole distribution of returns that accounts not only for the central tendency and dispersion of an asset, but for the whole aggregated behaviour including asymmetry, kurtosis and the tails. Methods to analyse distributional data belong to the field of symbolic data analysis (Noirhomme-Fraiture and Brito (2011)), where observations account for internal variation that can be represented as intervals or distributions, and have been previously used in finance (Arroyo and Maté (2009); Arroyo et al. (2011); González-Rivera and Arroyo (2012)). Finally, we consider the observed data, that is the log return time series that accounts for variations over time and makes possible to identify when volatile or stable periods take place in each cryptocurrency.

There is a high diversity of clustering techniques, but in our case the interest lies on the different perspectives shown at each level of granularity. Thus, for all the representations we will use a partitional prototype-based clustering algorithms with a similarity measure (distance) meaningful for each kind of representation. In this way, we will have a prototype describing the behaviour of each cluster using the same representation of the data. Prototypes make possible to assign a financial meaning to the whole cluster.

Then, we will combine the three clustering results and analyse the most numerous intersections with the help of visual tools. Such approaches are successfully used in biostatistics (Kern et al. (2017); L'Yi et al. (2015)). In our case, we will use them to represent the main trends in the cryptocurrency market. If several cryptocurrencies belong to the same clusters in the three clustering results then we can consider them as very similar. We will inspect the relationship among the three clustering results with the help of visualization tools.

This approach provides a screening mechanism that allows us a meaningful exploration of the whole market in spite of its complexity and size. The intersection

of the clustering results can also be helpful for investors in order to select a suitable cryptocurrency for the portfolio as it characterizes the cryptocurrency market in more detail.

In a further step, we will investigate the association between the clustering results and different features of the cryptocurrencies such as technological variables, the market capitalization, the *maturity* (age) of the cryptocurrency, and some of asset portfolio ratios. We aim to inspect whether some clusters are tightly associated or not with some aspects not taken into account the clustering process. we apply some inference statistical tests to assess whether associations are significant. These associations enhance the profiling of the different clusters. We keep continuous references to concrete cryptocurrencies of the market, most of them no very known, that are part of our analysis. Finally, we present our conclusions with some points open to debate.

Related work

Clustering financial data

Clustering analysis is a well-known data analysis tool that have served well in different fields ([Henning et al. \(2016\)](#)). In particular in Finance, the seminal work of [Mantegna \(1999\)](#) used the cross-correlation of the return time series and Minimum Spanning Trees (MST) to group the stocks of the New York Stock Exchange from 1989 to 1995. In [Mantegna \(1999\)](#) is applied the MST to represent the stock market as a network. Later, in [Bonanno et al. \(2004\)](#) applies the same methodology considering different time horizons comparing the return and volatility networks. The methodology of Mantegna is applied with different variations in other contexts ([Brida and Risso \(2009\)](#); [Mizuno et al. \(2006\)](#); [Onnela et al. \(2003\)](#)). Furthermore, different alternatives and variants of this methodology have been proposed as it can be seen in the review in [Marti et al. \(2017\)](#)

Another strand of clustering applied partitional methods as we do but more specifically fuzzy-types to financial markets typically used for grouping stocks for developing portfolios. In [Nanda et al. \(2010\)](#) applies K-means, Fuzzy C-means and Self Organizing Maps (SOM) to returns and financial ratios from Indian stocks to classify them in different clusters and subsequently develop portfolios from these cluster. A derivate approach is shown in [Chaudhuri and Ghosh \(2015\)](#), who group the daily Indian market volatility comparing Kernel K-means, SOM and Gaussian clustering models to achieve right volatility prediction using the clusters as predictors. In the same way, [Pierpaolo et al. \(2013\)](#), [D'Urso et al. \(2016\)](#) and [D'Urso et al. \(2020\)](#) applied different variations of fuzzy clusters to financial markets. Our methodology requires fully distinguished classes for the clustering objects, partial membership of *fuzzy* clusters is more complicated to handle with our approach so a pure crisp partitions is preferred for the methodology proposed.

In [Liao \(2007\)](#); [Liao and Chou \(2013\)](#) do clustering on the daily market data and apply different association rules between the K-means groups, indices and market categories. That associations help to analyse and describe the co-movement among the different markets.

Regarding the use of time series as objects to cluster, in [Aghabozorgi and Teh \(2014\)](#) propose a three-phase clustering model to categorize companies based on the

similarity in the shape of their stock markets using Dynamic Time Warping (DTW) (Berndt and Clifford (1994)).

From traditional finance to cryptoasset markets

In Yermack (2013) analyses Bitcoin market in-depth and considered it an investment more speculative than a currency. It is considered that it poses high risk for the management of transactions and credit markets. Finally, a deflationary scenario is anticipated because of the limited number of bitcoins that can be issued (21 millions). The paper anticipated many aspects of the cryptocurrency markets that we are experiencing today (excessive volatility, high level of computer knowledge required for using and integration into the web of international payments).

The high growth of the cryptocurrency market and its heterogeneity since 2014 was analysed in depth in Corbet et al. (2019) considering different aspects including regulatory, cyber-criminality, market efficiency or bubble dynamics and make recommendations for further investigations on different domains. We take a couple of them and we address in our work some characteristics based on liquidity (market cap) and other key market metrics, for instance, Beta or Modigliani-Modigliani financial ratios.

The characterization of the cryptocurrencies from a statistical point of view has been tackled by different works. In Chan et al. (2017) analyses the distributions for a few cryptocurrencies (Bitcoin, Dash, Dogecoin, Litecoin, MaidSafeCoin, Monero and Ripple) and show that they exhibit heavy-tailed distributions that fit the generalized hyperbolic distributions. As part of a benchmark with other markets, in Baek and Elbeck (2015) shows that the volatility of Bitcoin market shows that bitcoins are 26 times more volatile than S&P 500 Index.

In Zhang et al. (2018) analyses the stylized facts of eight cryptocurrencies that represent almost 70% of the market capitalization and find, among other things, heavy tails for the returns, return autocorrelations that decay quickly, while the autocorrelations for absolute returns decay slowly, that returns display strong volatility clustering and leverage effects, and a power-law correlation between price and volume. The study of stylized facts have been extended increasing the number of digital coins up to 222 Hu et al. (2019).

Clustering of cryptocurrencies

The classical methodology based on MST algorithms (Mantegna (1999)) is applied in Song et al. (2019) to filtering the influence of *Bitcoins* and *Ethereum* and detects six homogeneous clusters. However, the structure found does not remain stable after the announcement of regulations from various countries. Interestingly, the uses of clustering together with other methods, such as VAR models and Granger causality tests (Zieba et al. (2019)) help to find that Bitcoin shock prices are not transmitted to the prices of other cryptocurrencies, being *Litecoin* and *Dogecoin* the more influential actors. According to the results, Bitcoin exhibits a lower relationship with other cryptocurrencies. Other approach is the use of random matrix theory and hierarchical structures in a MST on 119 cryptocurrencies from 2016 to 2018 (Stosic et al. (2018)). It is showed that the presence of multiple collective behaviours in the market of cryptocurrencies, which contrast to the intuitive idea that Bitcoin exerts a global influence on the entire market.

Furthermore, the time dimension can also be taken into account evolve. In Sigaki et al. (2019) first classify 437 cryptocurrencies according to information efficiency using permutation entropy and statistical complexity, and then cluster their time series using dynamic time warping and hierarchical clustering to find four groups where the behaviour in terms of information efficiency evolves differently. All these articles evidence the complexity of the underlying structure in the cryptocurrency market, where some cryptocurrencies influence others even in unexpected ways.

The comparative study of cryptocurrency markets and traditional financial markets is also an essential domain of investigations. In Corbet et al. (2018) shows that cryptocurrencies are highly connected among themselves and disconnected from mainstream assets (bonds, stocks, S&P500, gold). In Pele et al. (2020) merge classification based on asset profiles and dynamic evolution of clusters. It starts with a characterization of selected group of log-returns assets including 150 cryptocurrencies, stocks commodities and exchange rates to estimate a multidimensional vector applying a dimension reduction (Factor analysis). Then, they use classification where K-means is one of the techniques applied. The main difference of Cryptocurrencies with respect to traditional assets is a higher variance and longer tails of the log-return's distribution. The work also shows that individual cryptocurrencies tend to develop over time similar characteristics (synchronic evolution). Drozd et al. (2019) shows that the multiscaling characteristics of the exchange rate fluctuations comparing the cryptocurrency market approach to those of the Forex; furthermore Begusic et al. (2018) estimates the scaling exponent and find the asymptotic power-law behaviour for Bitcoin.

Methodology

Dataset preparation

We retrieved data from <https://www.cryptocompare.com/> for all the cryptocurrencies traded during 2018. New cryptocurrencies have appeared in the last years, but many of them were short-lived and barely traded. We aim to incorporate in our descriptive study as many cryptocurrencies as possible on the market during the period which means that we are very conservative on the *clearing data* criteria, we only removed NaN and Inf observations mainly caused by zero-prices in log transformations. As result, from all the downloaded cryptocurrencies we filtered out those that were in the market less than the 95% of the days (92 cryptocurrencies in 2018), but we kept some that were in the market but were not traded, i.e. zero return and volatility or zero volume. Our final data set consists of 1,723 cryptocurrencies. Anyway, part of the cryptocurrencies present a very low activity in the market point to low liquidity that cause heavy-tails effect so for that case we apply a specific *filter* in some parts of the study, mainly for association tests

For these cryptocurrencies we constructed the following variables:

- **Daily log-returns:** The use of returns instead of prices in Finance price time-series is very extended and consolidated due to its more suitable statistical properties and better comparability. It has been used in cryptocurrency markets as well Letra (2016); Stosic et al. (2018). The return for the cryptocurrency i at day ts is computed as:

$$r_i(t) = \ln(P_i(t)) - \ln(P_i(t-1))$$

where $P_i(t)$ is the daily cryptocurrency price for i cryptoasset at day t .

- **Volume:** It is the daily traded volume in units of the base cryptocurrency that can be understood as a liquidity proxy. We transformed it into an ordinal variable by the quantile functions. Three cryptocurrencies represent 66% of the trading volume of the market in 2018, namely Bitcoin (46%), Ethereum (16.5%) and EOS (4%), and in total 10 cryptocurrencies (BTC, ETH, EOS, BCH, XRP, LTC, ICX, HSR, ETC, IOT) represent 80% of the daily volume.
- **Market cap:** it is the one-day market cap retrieved directly from the market on the 4th of February in 2019. Three cryptocurrencies represent 60% of the *market cap*, namely WBTC* (26.8%), BTC(22.4%) and NPC (11.5%), and in total 5 cryptocurrencies (WBTC*, BTC, NPC, XRP, AMIS) represent 80% of the *market cap*
- **Technological variables:** We represent the encryption and consensus algorithms of the cryptocurrency as nominal variables:
 - **Encryption:** There are 105 different values. The more relevant encryption algorithms are Scrypt, SHA256, SHA256D, X11, X13, X15, PoS, Multiple and CryptoNight. We notice that this information is not available for 35% of the cryptocurrencies (599 obs.).
 - **Consensus:** There are 60 possible values, including the well known Proof of work (PoW) and Proof of Stake (PoS). The most predominant are obviously PoW/PoS, PoW and PoS but this information is missing in 31% of the cryptocurrencies (536 obs.)
- **Age:** We estimate the time on the market of each cryptocurrency, and transform it into an ordinal variable by a quantile function. *Age* and *maturity* are interchangeable on our study.

Besides the cryptocurrency data, we also retrieve time-series from <https://cci30.com/> which is a rules-based index designed to objectively measure the overall growth, daily and long-term movement of the blockchain sector and the US Department of the Treasury^[3] that we use for the computation of some financial benchmarking rates as Beta, Sharpe ratio and others that we will explain in the next sections.

Volume of negotiation and tail distribution analysis

We aim to consider all cryptocurrencies that existed on the market independently of the transaction volume or liquidity. For sure liquidity or volume of trading is a key aspect to keep in mind to characterize the market but our first priority is to represent and manage all the market. We represent a histogram of non-trading days based on the volume negotiated (volume equal to zero means in our case no traded asset) on Figure 1 where 306 cryptocurrencies has been traded (volume of negotiation equal to zero) no day during 2018 (18 crypto currencies traded only 2 days). As we commented, we consider relevant to keep non-traded cryptocurrencies as part of our market analysis for the clustering part just to detect where they are allocated. However, we agree on the low interest for the investor in lower liquidity crypto assets.

[3] <https://home.treasury.gov/>

The concept of power-law distribution [Newman \(2005\)](#) and applications on cryptocurrency case [Begusic et al. \(2018\)](#), [Watorek et al. \(2020\)](#) is relevant to confirm the existence of 1st and 2nd order moments on our data to ensure the applicability and generalization of the results of clustering and the inference tests. The existence of the statistic moments according to the power-law [Gillespie \(2015\)](#) is:

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}, \quad (1)$$

and for the CDF

$$P(X \leq x) = 1 - \left(\frac{x}{x_{min}} \right)^{-\alpha+1}, \quad (2)$$

The moments of the power law distribution for the continuous power law are the following:

$$E[X^m] = \int_{x_{min}}^{\infty} x^m p(x) \cdot dx = \frac{\alpha - 1}{\alpha - 1 - m} x_{min}^m \quad (3)$$

Generally speaking when

- $1 \leq \alpha \leq 2$, all moments diverge, i.e., $E[X] = \infty$;
- $2 \leq \alpha \leq 3$, all second and higher-order moments diverge, i.e., $E[X^2] = \infty$;
- $3 \leq \alpha \leq m + 1$, all m and higher-order moments diverge, i.e., $E[X^m] = \infty$;

[Newman \(2005\)](#) showed that for the case $2 \leq \alpha \leq 3$ as X_{max} becomes large we get that

$$E(x^2) \sim n^{(3-\alpha)/(\alpha-1)} \quad (4)$$

converge. In [14](#), [15](#), [16](#), [17](#) and [18](#) we have computed the maximum likelihood estimator (MLE) for the continuous power law for the negative and positive tails on the variables **AlphaP** and **AlphaN** as follows:

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \log \frac{x_i}{x_{min}} \right]^{-1} \quad (5)$$

with the following expected statistical error (variables **Sd.P**, **Sd.N** for positive and negative tails on the table [14](#))

$$\sigma = \sqrt{n} \left[\sum_{i=1}^n \log \frac{x_i}{x_{min}} \right]^{-1} = \frac{\alpha - 1}{\sqrt{n}} \quad (6)$$

In our view, we consider that power-law behaviour could be more impacting on the generalization of the inference tests than clustering partitions. Our goal is to represent all the cryptocurrency market on a clustering maps independently of any statistical property or behaviour considering them as objects and in that way we

keep the 1723 cryptocurrencies for clustering. In other words, we want to know where are allocated the highest *heavy-tail* cryptocurrencies as well. For association tests our goal is the generalization of the results and in that way we filtered out the list of cryptocurrencies keeping only the higher statistic quality, 1262 cryptocurrencies in total.

Methods

We aim to find groups of cryptocurrencies based on the behaviour of their log-returns in 2018 and describe them. For such purpose we will use different clustering algorithms that will deal with the three representations of the log-returns described in the previous section: statistic moments, observed probability distribution and observed daily time-series.

We use centroid-based clustering algorithms because the centroids provide us an interpretable summary of the elements of each cluster, which will help us to identify the most relevant features of the cluster elements. A drawback in this type of clustering algorithms is that they assume knowledge about the desired number of clusters (k). We apply different quality criteria to determine the optimum number of clusters depending on the technique.

Moreover, we use distance-based clustering algorithms they are simple, intuitive and applicable for a wide variety of scenarios [Aggarwal et al. \(2013\)](#). The algorithms considered will be based on meaningful dissimilarity measures or distances that help on the interpretability of the clusters. That is especially important for more complex representations such as the distributions or the time series. For example, in the case of distributions, the measure should relate with properties of the density function (central tendency, spread, symmetry), while in the case of time series will be more with the shape of the time series along time. Meaningful measures will help us to understand better the resulting clusters and to interpret the nearness of the observations to the centroid. In addition, our clustering algorithm provide a prototype or a centroid of the clustering, which eases the characterization of the resulting clusters.

The cluster intersections help us to merge the results of the different clusters and identify the most prominent cryptocurrency profiles along 2018 according to different characteristics through the three techniques. Furthermore, we analyse the association between the clustering results found for the three representations and the different attributes of the cryptocurrencies.

The k-means clustering algorithm for the first and second statistical moments

For the bi-variate (or two-moments) representation, where the two variables are the yearly mean and standard deviation of the log-returns we use the k-means clustering [MacQueen \(1967\)](#), which is one of the most widely used clustering algorithms [Wu et al. \(2008\)](#). We standardize the two variables to homogenize the differences between their ranges. The K-means clustering minimizes within-cluster variances, that is, squared Euclidean distances in our case, which makes the result easy to understand and interpret. Before the clustering, we compute the Hopkins statistic [Banerjee and Dave \(2004\)](#) to rule out the possibility that a uniform random distribution generated the data set.

For the selection of the number of clusters (k) we compute several internal Cluster Validity Indices (CVIs) for crisp partitions [Arbelaitz et al. \(2013\)](#), including silhouette, Dunn, COP Davies-Bouldin, Calinski-Harabasz or the score function, and then apply the majority rule to choose the best number of clusters.

We apply clustering ensemble techniques [Acharya \(2011\)](#) aiming at reducing the randomness on the partitional cluster results. We run the K-means algorithms 10 times and we ensemble the outcomes minimizing the Euclidean distance. We confirm that the dissimilarity among the different runs is closer to zero, which makes the ensemble cluster a more stable representation. For each algorithm run, we apply the Hartigan-Wong method for clustering [Hartigan and Wong \(1979\)](#) with ten iterations to reach convergence and considering 50 random starts for each iteration. Once we have the 10 algorithm runs, we compute the medoid of an ensemble of partitions, i.e, the element of the ensemble minimizing the sum of dissimilarities to all other elements [Hornik \(2005, 2019\)](#).

The dynamic histogram clustering for the log-return distribution

For the yearly log-return distribution, we apply a clustering algorithm that deals with histogram-data form. More precisely, we apply the dynamic clustering algorithm for histogram data based on the l_2 Wasserstein distance [Irpino and Verde \(2006\)](#); [Irpino et al. \(2014\)](#). In this way, we will group the cryptocurrencies with similar distributions of log-returns in 2018.

The dynamic clustering algorithm needs a dissimilarity function to assign the observations to the clusters, which is the l_2 Wasserstein distance. Given two histograms h_1 and h_2 , the l_2 Wasserstein distance is defined as

$$d_W(h_1, h_2) := \sqrt{\int_0^1 [F_1^{-1}(t) - F_2^{-1}(t)]^2 dt} \quad (7)$$

where F_1^{-1} and F_2^{-1} are the inverse of the cumulative distribution functions, that's the quantile functions of h_1 and h_2 , respectively. This distance can be decomposed as follows:

$$d_W(h_1, h_2) = \sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2 + 2\sigma_1\sigma_2(1 - \rho_{1,2})} \quad (8)$$

where μ_i and σ_i are respectively the mean and the standard deviation of the h_i and $\rho_{1,2}$ is the correlation of h_1 and h_2 [Irpino and Verde \(2015\)](#). As a result, the l_2 Wasserstein distance can be decomposed in the addition of three elements that account for the histogram differences in terms of location, spread and shape, respectively. Interestingly, this distance matches the perceptual similarity that the human observes when comparing distributions [Arroyo and Maté \(2009\)](#). All these aspects make it a suitable distance for clustering distributions and, in our case, log-return distributions.

The Dynamic Clustering Algorithm for histogram data based on the Wasserstein distance (Hist DAWass) is a k-means-like algorithm for clustering a set of observations described by histogram variables [Irpino and Verde \(2006\)](#); [Irpino et al. \(2014\)](#).

Each of the k clusters is represented by a centroid or prototype and observations are assigned to the closest prototype. The prototype is the average histogram of the observed histograms for each variable. In our case, observations are described by a single histogram variable representing the distribution of log-returns and the resulting prototype is a histogram that averages the histograms of the observations that belong to the cluster [Irpino and Verde \(2015\)](#). As a result, the prototypes can be interpreted in a financial context as log-return distributions.

We use the clustering implementation in the R-package HistDAWass [Irpino \(2016\)](#). This implementation provides quality measure that is the percentage of Sum of Squared (SS) deviation explained by the model running the algorithm several times for each k . We run the clustering algorithm 20 times for each k and the solution is the best one among the repetitions, that is, the one that maximizes the SS.

The TADPole clustering for the log-return time-series

Time-series clustering is a challenging domain for clustering due to the high dimensionality of the objects and how they are ordered. As a result, many approaches have been proposed over time [Aghabozorgi et al. \(2015\)](#); [Liao \(2005\)](#); [Rani and Sikka \(2012\)](#).

We wish to cluster the time-series with similar volatility patterns in the same periods. For this purpose, Euclidean distance may fail to produce an intuitively correct measure of similarity between two time series, because it is very sensitive to small distortions in the time axis. However, other measures, such as Dynamic Time Warping (DTW), cope with this problem by *warping* non-linearly the time dimension to estimate their similarity. Nowadays, DTW is considered one of the most popular and useful shape-based measures [Aghabozorgi et al. \(2015\)](#).

However, DTW is intrinsically slow because of its quadratic time complexity, which hampers its applicability in clustering. Thus, we use the TADPole (Time-series Anytime Density Peak) [Begum et al. \(2015\)](#) clustering algorithm that extends the Density Peak (DP) clustering framework [Rodriguez and Laio \(2014\)](#) and exploits the upper and lower bounds of DTW to prune unnecessary distance computations which speed-up the convergence of the algorithm. As a result, TADPole produces a right answer quickly, and then refines it until it converges to the exact answer. Besides, the clustering algorithm only needs two parameters which makes it easy to use. Firstly a cut-off distance that define the thresholds to select the series and we set it to 2; and secondly, a windows-size that define the time frame to make the comparison between the series that we set to 3. Optionally we can also select the number of clusters (k) or we can let the algorithm to choose the optimal one based on the local density of points (closer series at some time based on some cut-off distance) using a “knee point finding” algorithm where points with higher values of $\rho_i \cdot \delta_i$, where ρ_i referred to the local density and δ_i is the distance from points with higher local density.

We will consider a different number of clusters k and compute the internal Cluster Validity Index (CVI) for each of them. As this clustering algorithm uses three distances, we used as CVI the Calinski-Harabasz index that secure the convergence of the algorithm for asymmetric distance measure.

TADPole allows to cluster time-series with arbitrary shapes which is very useful in our case because of the heterogeneity of the cryptocurrency market. In contrast,

TADPole clusters cannot be represented as "balls" in a metric plane as in K-means for example. The result is a Partition Around Medoid (PAM) type centroid using the DTW distance that can be represented in a DTW space. This centroid is a time-series that serves us to identify the volatility patterns of the resulting clusters.

We apply the R implementation of TADPole algorithm in the R-libraries [Sardá-Espinosa \(2019\)](#); [Sardá-Espinosa \(2019\)](#). The time series is on log-return values which facilitate the characterization of the clusters. The DTW measure implemented in the package follows the estimation by [Lemire \(2008\)](#).

Combination of clustering results

Once we have the results of the clustering algorithms, we combine them by intersecting the clusters. Potentially we have $T_1 \times T_2 \times T_3$ intersections where T_n is the number of clusters that we obtain with the clustering algorithm n . The combination of the clustering results would make possible to characterize each cryptocurrency in several dimensions, one for each cluster strategy. The resulting multi-dimensional categorical datasets can be shown using visualization techniques supported on the graph theory [Kern et al. \(2017\)](#); [L'Yi et al. \(2015\)](#). To better highlight the changes in the clustering between the different techniques, we have visualized such changes by means of a so called alluvial diagram [Rosvall and Bergstrom \(2010\)](#). We use the alluvial visualization implemented in R-library [Bojanowski and Edwards \(2016\)](#) to show the main *flows* of cryptocurrencies.

We will analyse the clustering intersections with the higher cardinality to better profile the main trends of the cryptocurrency market.

Association test

Finally, we complement the information of each cluster result by studying the level of association with different independent variables not considered in the clustering process. We analyze the association among clusters and the categorical variables defined in Table 1 by applying the exact Fisher's tests and analysing the Person's residuals of the contingency tables that we explain later. Firstly, we transform quantitative variables into ordinal by applying the quantile functions.

We introduce below some variables that are financial ratios [Bacon \(2008\)](#) that we borrow from portfolio theory and we use to characterize the behaviour of the cryptocurrencies from a investor perspective, enhancing the association study as well. The computation of most of the below ratios (except Beta that has been manually computed) rely on the ratios implemented on R-library [Peterson et al. \(2018\)](#).

Beta is a volatility measure of systematic risk of an asset, the risk inherent to the entire market that is non-diversifiable, in statistic terms, *beta* is the slope of the regression of our asset compared with a reference on the market:

$$\beta = \frac{Cov(R_c, R_b)}{Var(R_b)}, \quad (9)$$

where R_c is the return of our cryptocurrency, R_b is the return of the bench market, the CCI30 index that track the 30 largest cryptocurrencies by market capitalization,

excluding stablecoins. . The *beta* value shows whether an asset moves in the same direction as the reference index, and how volatile or risky is compared with it. The Beta for the whole market is 1.0. A positive beta means the asset moves in the same direction as the market, while negative beta means that the asset moves in opposite direction. Furthermore, an absolute value greater than 1 means greater sensitivity to systematic risk, i.e. higher risk; while values lower than 1 mean less sensitivity.

The **Sharpe ratio**(Sharpe variable) is the exceed average return of risk-free by volatility unit or total risk. The ratio determines the risk of the investment with respect to the return of an investment with zero-risk:

$$SR_c = \frac{E[R_c - R_f]}{\sigma_c}, \quad (10)$$

where R_c is the return of our cryptocurrency, σ_c is standard deviation or the volatility of our cryptocurrency and R_f is the *risk-free* rate taken as reference; we considered the daily of the annualized T-Bill over 90 days, and its daily value for 2018 was $E[R_f] = 0,005254377\%$. The greater the value of the Sharpe ratio, the more attractive the risk-adjusted return of the cryptocurrency.

Typically, **Chi-Square test** is used to examine the significance of the association between categorical data on a contingency table. However, the significance value is an approximation that it is not adequate when the sample size is small. We ruled out the Chi-Square test since results are not significant if the expected frequency is not typically higher than 5 in at least 80% of the cells of the contingency table (Yates (1984)) and this assumption is not fulfilled in our case for many of the categorical variables for some levels. we will use the Fisher's exact test (Fisher (1922)) to test the association between the variables of the Table 1 and the cluster results, which is applicable for all sample sizes. This test assumes no dependency between the categorical variables as null hypothesis, and assumes a multivariate hyper-geometric distribution for the cells into the contingency tables (Mehta and Patel (1983)).

For large datasets Monte Carlo method provides an unbiased estimate of the exact *p-value* (Mehta and Patel (1996)). Monte Carlo is a repeated sampling method that for any observed table, there are many tables, each with the same dimensions and columns and row margins as the observed table. Monte Carlo simulations are implemented in R stats-package for the *chisq.test* function. We run 8,000 simulations for each association, i.e. for each pair of variables under analysis, generating simulated contingency tables filled with a sampling of a multivariate hyper-geometric distribution. Then we compute the probability that we have a distribution as we effectively have observed, that is the *p-value*. A cell-by-cell comparison of observed and estimated frequencies evidences the nature of the dependence. If the p-values of the Fisher association tests between a couple of variables is lower than 0.01 then we consider that the association is significant. For each significant association between categorical variables of the contingency table we analyse standardized (adjusted) Person's residuals for the cell ij (Agresti (2018)), which is defined as follows

$$r_{(Adj)ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}(1 - \frac{m_i}{N})(1 - \frac{n_j}{N})}} \quad (11)$$

where O_{ij} and E_{ij} are the observed and expected frequency, respectively, m_i is the row total, n_j is the column total, and N is the total number of observations.

The sign of the residual (positive or negative) indicates whether the observed frequency in cell ij is higher or lower, respectively, than the value fitted under the model, while the magnitude indicates the degree of departure. A standardized residual having an absolute value that exceeds about 2 when there are a few cells, or about 3 when there are many cells indicates that the cell does not satisfy H_0 (Agresti (2018)). In our case, we assume a more conservative position and we consider as a cut-off for significant standardized residuals those that exceed 3.5.

We will use association plots to indicate deviations from a specified independence model of our contingency table. The association plots provide a quick means for visualizing the residuals of an independence model for a contingency table (Meyer et al. (2006, 2020); Zeileis et al. (2007)). In the association plot, each cell is represented by a rectangle that has (signed) height proportional to $r_{(Adj)ij}$ and width proportional to $\sqrt{E_{ij}}$, so the area of the box is proportional to the difference in observed and expected frequencies. The rectangles in each row are positioned relative to a baseline indicating independence ($r_{(Adj)ij} = 0$). If the observed frequency of a cell is greater than the expected one, the box rises above the baseline, and falls below otherwise. Additionally, the residuals can be coloured depending on a specified shading scheme. In our case, blue tones for positive residuals and red tones for negative ones, darker tones will be associated for the higher residual absolute values.

Results

In this section we present the results of the three clustering algorithms, of the intersection clustering and finally of the association tests. In Table 2 we summarize the three clustering results, showing for each cluster its cardinality and, for the sake of comparison, the observed mean and standard deviation of the prototypes (for the K-means we show the centroid values).

Clustering results of the bi-dimensional (one-two moments) representations

For the existence of clusters, the Hopkins statistic computed on scaled average returns and volatility is 0.01552. The value is below 0.5 which points out the existence of an underlying structure.

The optimum number of clusters according to the CVI indexes is three. The descriptive statistics of the three centroids in ordinary values are shown in Table 2 and Figure 2(a) shows the scatter plot with the clusters.

The algorithm clearly discriminates the cryptocurrencies between lower (Clusters 2 and 3) and higher volatility (Cluster 1) which is the less populated cluster as well. From a financial perspective, Cluster 1 includes the riskier cryptocurrencies. The cryptocurrencies of Cluster 3 mostly have negative mean returns, while those in Cluster 2 have higher returns, some of them positive and others negative. In general, the three centroids are close to the zero mean return point.

Figures 3(a), 4 and 5 show a more detailed view of each cluster. In these figures we represent in a contour plots the density in the bi-dimensional (two-moments)

space (\bar{r}, σ) that help us to locate the areas where cryptocurrencies tend to be more concentrated.

- **Cluster 1:** This cluster allocate cryptocurrencies with negative average returns, but with very high volatility, ranging from 1 to 5. It includes only 19 cryptocurrencies that represent around 1% of the sample as we see in Fig. 3(a). The higher concentration of cryptocurrencies into this cluster is surrounding volatility 1.5 and mean return around -0.1 as we see in Fig. 3(b). ELTCOIN (token that run on Ethereum blockchain network released in October 2017) has a central position in the cluster and around it we can find B2X, ADCN (no traded on the market since November 2019), BLX, WAND (derivative market platform), GOOD, SBIT, ZCG, ITT, REX, STAR (it is a token and operates on the Ethereum platform, higher volume in Ethereum along 1st and last quarter of 2018) and PFR. We see into this cluster a mix of Ethereum tokens and cryptocurrencies with its own blockchain, most of them with low traded volume which may cause that a few operations trigger the volatility.
- **Cluster 2:** This cluster is the more populated with around 900 cryptocurrencies (52% of the total). It allocates the *moderate* behaviours including the higher mean return cryptoassets. It is also less homogeneous than the others, with different dense areas of concentration as we see in Fig. 4(a) which point out the existence of other cluster. Most of the high capitalization cryptocurrencies (BTC, EOS, ETC, ETH or LTC) are in the sub-cluster with very low volatility and moderate negative return shown in Fig. 4(b). However, we can also find some cryptocurrencies with moderate positive returns (134 cryptocurrencies) and very low volatility, as shown in Fig. 4(c). These cryptocurrencies include ALEX (low trading in the first half of 2018 and higher activity in the second half of 2018), BST (BlockStamp had very low activity along 2018), ETL (EtherLite is a ERC20 token based on Ethereum with high peaks of activities in the first quarter of 2018 and no activity in the remaining part of the year) or OPES (OpesCoin had a moderate activity in the first half of 2018, and was flat in the second); all of them can be considered low-medium market capitalization (under 70th percentile). In Cluster 3, it is also possible to find a few cryptocurrencies with higher returns, but they have a low market capitalization.
- **Cluster 3:** This cluster has 801 cryptocurrencies, most of them with negative average returns, and volatility lower than 0.5. According to Figure 5(a), the highest concentration of cryptocurrencies is located in mean return closer to zero and volatility around 0.1. Some of the more representative cryptocurrencies of this cluster in terms of market capitalization are XEM, VIA, QRL, DASH, QTUM, XST and BCH that are close in the cluster (see Figure 5(b)).

We confirm that the K-means clustering identifies three different behaviours on the cryptocurrencies in terms of mean returns and volatility.

Clustering results of log-return histogram representations

According to the CVI, the clustering algorithm for histogram data based on l_2 Wasserstein metric [Irpino \(2019\)](#) separates the cryptocurrencies into five clusters. Each cluster is represented by its prototype, which is a log-return distribution. Table 3 shows the descriptive statistics from the prototypes of the five clusters.

The five distributions exhibit a slightly negative central tendency measures, being Cluster 1 the one with the lowest values. They are quite symmetrical, with low skewness and heavy tails pointed out by a high kurtosis. Skewness is closer to zero in all cases but it is positive, which means that the right tail of the distribution is fatter, or in other words, it has more extreme positive return values (or over the mean) on the right.

It is important to remark that the coefficient of variation for the centroids is quite different in all the clusters ranging from -0.75 to -32.90, which point out that this clustering algorithm is specially sensitive to this particular statistic. This is particularly relevant in this financial context, since the coefficient of variation measures how much volatility, is assumed in comparison to the amount of return expected from investments. However, since the mean returns are negative its financial interpretation would be misleading.

The last column of Table 3 provides a variance measure that quantifies the deviation of the distributions of a cluster with respect to its prototype. It is a dispersion measure for histogram data based on the L2 Wasserstein metric [Irpino and Verde \(2015\)](#). This statistic measures how much representative is the prototype of a cluster. According to this statistic, Cluster 3 would be the more uniform cluster, while Cluster 5 would be the more heterogeneous.

The first column of Figure 6 shows the prototypes of the five clusters in each row, while the rest of the columns show relevant cryptocurrencies from each cluster. It is interesting that except for the prototype of Cluster 1, the others exhibit a similar shape, where the main differences lie in the range of the distribution (note that each plot has a different range for the X axis) and in the tail behaviour. We will describe them below:

- **Cluster 1:** The prototype shown in Fig. 6(a) has a mean return of -0.13 and the highest kurtosis (13.43). The standard deviation of this prototype is slightly lower than that from Cluster 1 prototype, however, the shape of the distribution is different, because here the tails are thicker. The Wasserstein variance associated with the mean distribution (0.025) suggests that the cluster is homogeneous and has a cardinality of close to 500 cryptocurrencies that represent around at 30% of the sample. Some of the most representative cryptocurrencies in this cluster have a high market cap (P99), for example, BITUSD (high capitalization along 2018 but in a downward trend), CHAT in Fig.6(b), KEY in Fig.6(c) (high trading volume in the second half of 2018), MAN (increasing trading volume along 2018, maximum at the end of the year) and OCN (a token for peer-to-peer sharing economies such as Airbnb).
- **Cluster 2:** The prototype shown in Fig. 6(d) (green color distribution) has the lowest mean (-0.50) and median (-0.51) return among the clusters, and the highest coefficient of variation (-0.75). The cluster variance (0.079) points out a quite homogeneous cluster. This cluster has a cardinality of 147 cryptocurrencies and concentrate all no-traded cryptocurrencies (92), lowest market cap (P70) for most of the cryptocurrencies into this cluster. Representative into the cluster by market cap are 365 in Fig.6(e), ACN, CBX or ALT in Fig.6(f).
- **Cluster 3:** The prototype shown in Fig. 6(g) has a mean return close to zero (-0.01) and the most moderated volatility (0.11) and the shortest observed

range between minimum and maximum returns. According to the Wasserstein variance, this cluster is the most homogeneous, which is specially interesting given that it has the highest cardinality with more than 1000 cryptocurrencies (around 60% of the sample). Unsurprisingly, this cluster allocates the cryptocurrencies with the highest market capitalization, including BTC in Fig.6(h), BCH, EOS, ETC, ETH in Fig.6(i) and others (HSR, ICX or LTC). Given the size of the cluster, these cryptocurrencies represent the predominant behavior in the market, which unsurprisingly is the most moderate behaviour and includes the most popular cryptocurrencies.

- **Cluster 4:** The prototype shown in Fig. 6(j) is characterized by negative mean returns (-0.04), notable volatility (standard deviation of 0.87) and fat tails with very high kurtosis (11.95). The coefficient of variation is low too (-19.97). The cluster is not very homogeneous compared with the mean distribution (0.128). The cardinality of this cluster is low (around 60 cryptocurrencies) and some of the representative cryptocurrencies are NAS (most of the trading volume in 2nd and 3rd quarters of 2018), NKC (high trading volume since February 2018 and very important trading volume in Aug 18), POLY in Fig.6(k) (launched in January 2018), FSN (higher volume activity in 2nd and 3rd quarter of 2018 with a peak in August), JNT in Fig.6(l) (higher volume activity in 3rd quarter) or MNTP (no continuity on the trading volume with sporadic peaks).
- **Cluster 5:** The prototype shown in Fig. 6(m) has a mean returns closer to zero (-0.09) but the highest standard deviation (3.12), which causes the lowest coefficient of variation (-32.90). The shape of the cluster is almost symmetric (0.05) with a moderate kurtosis compared with the others clusters (5.66). We find the highest negative and positive returns in this cluster. This cluster is the most heterogeneous compared with the mean distribution (1.116). Unsurprisingly, it has the lowest cardinality with only 16 cryptocurrencies, which is around the 1% of the sample. Some of the representative cryptocurrencies are B2X in Fig.6(n) (low trading in 2018), ITT in Fig.6(o) (very active trading volume in January 2018 and along and a peak in July but no activity since that time, no trading volume in 2019), LBTC (launched in 2017, discontinuous activity along 2018 with no activity at all from September to end of November), PFR, STAR (noted in Cluster 2 of K-means), YOVI, AMIX, ELTCOIN (noted in Cluster 2 of K-means) or FLLW (some activity the first 2-3 months of 2018, low trading volume in the remaining part of the year).

The HistDAWass clustering shows that is possible to effectively discriminate the log-return distributions taking into account central tendency, dispersion and shape.

Clustering results of the log-return time-series representation

The TADPole clustering Begum et al. (2015) has better performance with a $k = 3$ value according to the Calinski-Harabasz index. Figure 7 represents the time series of the medoids of each cluster so they are observed objects (time series of cryptocurrencies). Figure 8 shows the annual and quarterly density functions of the three medoids.

- **Cluster 1:** The medoid of this cluster in Figure 8(b) shows a time-variation around zero with return peaks positive and negative up to $(-0.2, +0.2)$. The central part of the distribution is heavily concentrated around zero, but with extreme volatility. The quarterly average returns changes smoothly starting with a low but positive value the first quarter, negative the second and third, and positive the fourth. This cluster has the lowest cardinality (22 cryptoassets). The medoid of this cluster is the time-series LINK (Chainlink's native token, known as LINK, is used to pay the network's node operators, or oracles, for providing secure data feeds). Other cryptocurrencies in this cluster are LTCU, PPC, SWT, AIR, NGC, PLR or ZSC
- **Cluster 2:** The medoid of this cluster in Figure 8(c) shows a consistent average returns above zero. The density functions have three modes and they are greater than or equal to zero. However, the last two quarters of 2018 exhibit fat negative tails with ranges over -0.1 . The cardinality of this cluster represents around a 49% percentage of the cryptocurrencies and it includes some of the highest market cap cryptocurrencies BTC, HSR (noted in Cluster 3 of Hist DAWass), ICX (noted in Cluster 3 of Hist DAWass), LTC (noted in Cluster 3 of Hist DAWass) and XRP. The medoid is the cryptocurrency XTO (called Tao coin as well is a token for music streaming services).
- **Cluster 3:** The medoid of this cluster in Figure 8(d) has average returns below zero in all the quarters and the densities exhibit two modes smaller than or equal to zero and occasionally large positive returns. The cardinality of the cluster is around a 50% of the total. This cluster includes most of the remaining highest market cap cryptocurrencies as EOS, ETC, ETH. The medoid is the cryptocurrency ZNE (Zone coin with more trade activity in the first quarter of 2018 with the more important peak of trade volume in July, flat trading volume the remaining part of the year).

We can see that the TADPole clustering for the return time series effectively identifies three different clusters taking into account the time series trend and dispersion over time. In Table 4 we show the variability of the clusters, measuring the variability as the mean distance (DTW+LB) to the centroid and its standard deviation with LB as *Lower Bound*. The variability is quite similar in all the clusters, being cluster 1 the most homogeneous and cluster 3 the less. However, according to the standard deviation and the coefficient of variation the dispersion within the clusters is quite high.

Intersection of clusters

For the sake of comparison, Figure 2 shows the three clustering results on the same annual return-volatility plane. In each plot, all cryptocurrencies are site in the same location, but the colour scheme in each plot represents the respective clustering results. In the plot, we have marked the cryptocurrencies with highest market capitalization. We can see that most of them are very located in a precise area, below the point $(0,0)$.

The polygons and the colours reveal that the results of the three techniques are overlapped because respond to a different dimensionality on the objects. The only exception is the Cluster 1 of K-means in Fig.2(a) and Cluster 5 of HistDAWass

in Fig.2(b) which are mostly the same. These plots confirm that each clustering algorithm takes into account different aspects of the cryptocurrencies and that their combination may provide further insights on the cryptocurrency market. TADPole clustering in Fig.2(c) is the more different on the groups compared with previous techniques overlapping all the cluster areas for the same return-volatility plane.

We analyze now the main groups of cryptocurrencies that remains together through the three clustering algorithms, it is what we call intersection clusters. Only 24 out of 45 ($3 \times 5 \times 3$ intersections) possibles are populated. Table 5 shows the 6 most numerous, those with cardinality greater than 100, represents the 75% of total market.

Intersection 1 and 2 have almost 300 cryptocurrencies. Both of them are characterized by cryptocurrencies that belong to Cluster 2 and 3 in the K-means and HistDAWass algorithms, which unsurprisingly are the most populated clusters for each algorithm. Both of them are characterized by low volatility and (negative) close to zero average returns. However, in Intersection 1 we can find cluster 3 of the TADPole algorithm, while in Intersection 2, we can find cluster 2, which mainly differ in that in the first case it has negative quarterly average returns, while in the second case they are positive. In Intersection 1 we can find cryptocurrencies such as EOS, GVT, MANA, ETH or ETC. While in Intersection 2 we find some of the most popular highest market cap cryptocurrencies (BTC, LTC, XRP), and some others with lower market cap and higher returns (AE, USDT, ZRX).

Intersection 3 and 4 have around 200 cryptocurrencies with a high influence of K-means and Hist DAWass clusters. These intersections are characterized by cryptocurrencies that belong to Cluster 3 in the K-means and to Cluster 3 in the HistDAWass algorithm 6(g). The main difference with the previous intersections is that Cluster 3 of K-means corresponds on average with negative daily mean-returns but moderate volatility 3(a) so the average returns are lower as well for this intersection.

In the case of Intersection 3 we find one of the highest market cap cryptocurrency (BCH), and others with higher market capitalization (GNT, LSK, QTUM). While in the case of Intersection 4, the negative effect on average on the returns introduce by K-means Cluster 1 is compensated in some way by the TADPole Cluster2 with a centroid with all quarters over zero 8(c), no high returns cryptocurrencies on this intersection (DASH, SC, STRAT).

Finally, in Intersection 5 and 6, we have cluster 3 from the k-Means, cluster 1 from HistDAWass and clusters 3 and 2 from TADPole, respectively. Cluster 2 from HistDAWass was more volatile than cluster 3 and has heavier tails. In Intersection 5 we find cryptocurrencies with average-high risk and average returns (CMT, ETT, HST). Intersection 6 allocate some cryptocurrencies with high market cap but low returns (BCD, SBTC, GEO).

In the alluvial plot shown in Figure 9 we show how are related the different clusters of the different algorithms. It makes possible to appreciate both, the main trends already commented and those more subtle. For example, we can see that the smallest clusters in K-means and HistDAWass (Cluster 1 and Cluster 5, respectively) share most of the cryptocurrencies. In that sub-group of very volatile cryptocurrencies we find AMIS, B2X, ELTCOIN, FLLW, GOOD, ICE, ITT, LBTC, PFR, REX, RIPT, STAR, WAND, XIN, YOVI and ZCG. Then the group diverges and relates with the two main

clusters of the TADPole algorithm, which means that the temporal evolution is more conventional with mean return in quarterly basis positive or negative.

Interestingly, the smallest TADPole cluster (Cluster 1) is not heavily related with any other cluster. This means that its peculiar time series evolution as we see in Fig. 8(a) is not particularly related with the prototypes of the aggregated representations, namely as return distributions and mean-standard deviation bi-variate or two-moments representations.

Finally, we mention that DEUR is the only cryptocurrency that is not pair-combined with any other cryptocurrencies along the three techniques.

Heavy-tail cryptocurrencies on the clusters

Finally, we mention where are allocated the cryptocurrencies with more extreme behaviour, 461 in total. The Table 13 shows the distribution on the different clusters for each one of the techniques respect to the total cryptocurrencies allocated in that cluster. We conclude that heavy-tail distributions are more often allocated in Cluster 1 in K-means and Cluster 2 in Hist DAWass. We have already commented that Cluster 1 in K-means allocate the highest volatile cryptocurrencies and that correspond with heavy-tails cryptocurrencies as well. As we commented, Cluster 2 in Hist DAWass allocates the more negative-return cryptocurrencies so we conclude that heavy-tails are more strong in negative part of the distributions (or histograms for Hist DAWass clustering).

Association tests

As we explained in Methods sections, we rely on exact Fisher tests base on Monte Carlo simulations for the significance tests on the association between the variables. P-values of Fisher test are depicted graphically in Fig. 10 with the results of the Fisher tests among the categorical variables in Table 1 and the clusters (including the combination of clustering results). P-values lower than 0.01 are represented in purple colour addressing the more significant associations.

The goal of the association tests is to enhance the characterization of the clusters based on the prototypes that we explained in previous Section. We group in a red box the area with the associations between the clusters and the market categorical variables.

We represent in Figure 11 the mosaic plots portraying Person's residual of some relevant pair-variables. As we explained in Methods section, the height of the box is proportional to the corresponding Person's residuals representing the difference between observed and expected frequencies, while the width is proportional to the square root of the estimated expected counts. The colour intensity is proportional to the absolute value of the residual.

Association between market cap, volume and clustering results

According to Table 6, the Cluster 3 of K-means, we recap that it is the one with the prototype with the most pronounced negative mean returns, is associated with

the higher Volume percentiles (P80, P90 and P99) with standardized residuals 4.36, 4.35 and 5.71 respectively. However, Cluster 1 with the least pronounced negative mean returns, is associated with the lower percentiles (P70) with the highest residual 8.93. This is an interesting association since Volume was not considered as variable in the clustering process and shows an interesting and significant association for the behaviour of the cryptocurrencies that are not so popular with lower Volume or liquidity.

Finally, the association with the cluster for the highest market cap (*P100*) is not very strong but significant with a standardized residual of 3.71, most of the highest market cap cryptocurrencies are allocated in Cluster 2. For instance, some of the cryptocurrencies with highest capitalization (BTC, EOS, ETC, ETH, LTC and XRP) are located in Cluster 2.

According to Table 6 for Hist DAWass which shows the association for the case of the HistDAWass clustering, Clusters 1 and 2, whose prototypes had the lowest mean returns, are strongly associated with lower Volume cryptocurrencies with residuals 12.25 and 3.72, while Cluster 3, whose prototype had the least pronounced negative average returns and the lowest volatility, is associated with the highest percentiles with 7.85 as standardized residual por P99. It is also possible to see weaker associations in the smaller clusters (Cluster 4 and 5).

For the association between Market cap and Hist DAWass we confirm a high significance on the association between Cluster 1 and the lowest market cap percentiles with standardized residuals of 8.07. Cluster 3 groups medium-high market cap cryptocurrencies P90 and P99

Since we can find association in all the clusters, we confirm that the screening by market cap variable is equally significant for both techniques K-means and HistDAWass although the granularity is not too much fine separating the cryptocurrencies in two major groups, one for the lower market caps in one or two clusters and the remaining in another cluster.

Association between financial ratios and clustering results

Table 8 shows the strong association (standardized residual of 17.79) between the K-means Cluster 1 and the Extreme Beta values, consistent with the highest volatility of that cluster. Extreme Beta values (ICE, ITT, PFR and STAR) are associated with Cluster 1 with the more volatile cryptocurrencies (BLX, RIPT or XIN*). On the opposite, Cluster 2 that allocates cryptocurrencies with positive and moderate negative mean returns is associated with a high significance (standardized residual 6.70) with low volatility (*LowVol*) Betas (BTC, DCN, WAVES or WBTC*). Finally, Cluster 3 is associated (standardized residual of 5.79) with cryptocurrencies with high volatility Betas (*HighVol*) (ADA, BCH or SALT).

We confirm for Hist DAWass in Tab 8 the high screening capacity if this technique by separating with a high significance *NegBeta* in Cluster 1 and 2, *LowVol* with a

reasonable residual of 3.1 in Cluster 3, *IndexLike* in Cluster 3 as well and *Extreme* in Clusters 4 and 5 with the highest residuals (10.33 and 19.24)

Time-series clustering algorithm (TADPole) exhibits significant association with Sharpe ratio in 7

We can see that *ERP* (Excess Return Positive) category is associated with Cluster 2 (BTM, SC, DNT, LEND or WINGS) with very high standardised residual (10.60) and the *SRF* (Small Risk-Free) class in Cluster 3 (EOS, ETC, ETH, NEO or ZEC) with a high residual (11.65) as well. The higher quality Sharpe ratio Acc values are significantly associated with cluster 2.

Summarizing, the association of the clustering results, *K-means* and *Hist DAWass* clusters are associated with the market cap, volume and the Beta. Moreover, the TADPole clustering is associated with *Sharpe ratio*.

Association test for intersection of clusters

As we saw in Fig. 10, the cluster intersection (Combi variable) is significantly associated with most of the variables. The intersections provide a better characterization of the cryptocurrencies because of smaller groups, where it is easier to find stronger associations inherit of the combined clusters. 6, 8 and 7 show the association between different categoric variables and the more relevant intersections.

For **Volume**, Clusters 3 and 4 with standardized residual from 4.53 to 7.35 allocates higher liquidity cryptocurrencies with P90,P99 (ADT, BLOCK, CND) and even P80 percentiles (FLIX, LDC, RVT). The highest percentile P100 is mostly allocated in Intersection 1 (EOS, ETC, ETH) with 4.02 as standardized residual. Lower percentile P70 is mostly allocated in Intersections 1 (coincidence P100), 2 and 5, 6. Lower market cap cryptocurrencies are mostly allocated in Intersections 5 (ANTI, BBT, XMG), 6 (BTA, CNT, NTRN) with the higher standardized residuals. Medium cap are mostly allocated into the intersections 3 (ADT, BTX, ION) and 4 (BAY, LEND, SKY).

For the Sharpe ratio, the more acceptable for investment cryptocurrencies (Acc) are mostly allocated into the Intersection 2 (WAVES, XRP, ZEN). Exceed Return Positive (*ERP*) are divided into Intersections 2 (AC, ZRC, ZRX) and 4 (ADA, HSR, SC).

There is not any *Extreme Beta* for the higher cardinality Intersections as we see on the Table 8. We can see a good screening on the Beta values with high standardized residuals for Intersections 5 (FRX, PPP, XPY), 6 (GLC, TIT, XHI) for *NegBeta*, Intersection 2 (BTC, WAVES, WBTC*) for *LowVol* and Intersections 3 (BCH, QTUM, XVG) and 4 (ADA, HSR, STRAT) for *HighVol*. Finally, as expected, the combined cluster variable (Combi) inherit the associations of the different clustering techniques. and it is associated with all considered categorical variables except with the technological ones.

Associations for the technological variables

It is worth noting that while technological variables are not associated with any clustering results, they have significant relationship with the market cap (**PercMKCaP**) and volume (**PercVolume**) of trading (grey box). For example, in Table 12 there is a significant association between **Script** (7.58 standardized residual), **SHA256** (3.58) and **X11** (6.65) encryption algorithms and the percentile (P70) of market cap as well. Encrypted algorithms **CryptoNight-V7**, **Ethash**, **Ouroboros** seems more associated to the highest market cap percentile (P100).

Concerning the consensus algorithm (**ProofType** variable) in Table 11, **PoS** (3.74), **PoW/PoS** (9.09) and **PoW** (4.54) are clearly associated to the lower percentiles (P70) of market cap.

Associations with the age (maturity) of the cryptocurrencies

Table 9 and Figure 11 shows the association between **Age**, **K-means** and **HistDAWass** clustering results. More particularly, the Cluster 2 of the K-means, which was characterized by low volatility and negative and close-to-zero average returns, is associated with younger cryptocurrencies (**D4**). In the same way with the Cluster 1 of **HistDAWass** (**D4**), most of the younger cryptocurrencies of the Cluster 1 of **HistDAWass** remain together into the Cluster 2 of K-means. We remind in Table 2 that the cardinality of Cluster 2 of DAWass is 147 so 80% of the total cryptocurrencies in this Cluster are recent cryptocurrencies with the lower mean value for the returns 3 (-0.50) and medium volatility (0.38).

More mature cryptocurrencies (**D10**) are associated with **HistDAWass** Cluster 3 that are also the higher market cap clusters (**BTC**, **DASH**, **ETH**). Average maturity (**D5** (**ALIS**, **HDG**, **WTT**) and **D6** (**ATM***, **BTH**, **GRE**)) cryptocurrencies are associated into **HistDAWass** Cluster 1. In Figure , we see very high Person's residuals (up to 13) and a well-defined granularity by age for the **HistDAWass** clusters confirming this technique helpful for maturity screening.

Considering both clustering techniques, K-means and **HistDAWass** and the maturity, we confirm that lower maturity cryptocurrencies seem allocated in lower volatility and coefficient of variations clusters. We confirm an association between highest market cap, higher maturity, lower volatility 3 and higher liquidity in Cluster 3 of **HistDAWass** (**BCH**, **BTC**, **DASH**, **EOS**, **ETC**, **ETH**, **IOT**, **LINK**, **LTC**, **NEO**, **WAVES**, **XLM**, **XMR**, **XRP**, **ZEC**, **ZRX**).

For the intersection of clusters, Table 9 shows how the higher maturity (**D10**) cryptocurrencies are allocated into the Intersections 3 and 4 with very high standardized residual (7.20, 8.37 respectively). Average maturity (**D5**, **D6**) in Intersection 5 and 6, and younger cryptocurrencies (**D4**) are significantly allocated into the Intersection 6 (**EBC**, **ICOB**, **PULSE**).

Discussion

In this section, we summarise the main results obtained in the clustering and the association tests.

- We confirm the existence of a structure on the market that allow us to segment the cryptocurrencies in different clusters. The optimum number of clusters remains low, independently of the representation considered, which seems to point out a high degree of homogeneity despite the high number of cryptocurrencies.
- The bi-dimensional or two-moments representation by the classic K-means works quite well to segment separately by the returns and volatility in three major groups. However, the log-return distribution and the HistDAWass clustering offered a more subtle discrimination of the cryptocurrencies specially taking into account the combination of both moments, as we could see in the coefficient of variation values. Thus, it seems a more suitable profiling tool for investors.
- Both K-means and HistDAWass partitions split the data set into two groups according to market capitalization and Volume.
- The log-return time series representation provided a low number of clusters. However, they were significantly associated with Sharpe ratio by highest and lowest values.
- The K-means and the HistDAWass clustering provide an interesting association with the age or maturity of the cryptocurrencies. The results seem to point out that younger and older cryptocurrencies have particular and different return distribution.
- The intersection of clusters inherit by combining the association rules that we had observed separately for each one of the clustering algorithm. This is confirmed for Age, MkCap variables and the different financial ratios.

Conclusion

In this work we analysed the whole cryptocurrency market in 2018, that is all the cryptocurrencies traded in 2018, with a novel method that involved the combination of three different clustering algorithms. Each method used a meaningful representation considering different aggregation or granularity level of the daily returns: from the yearly average return and volatility, the yearly distribution of returns, and finally the observed time series of daily returns. For each representation we used prototype based clustering methods, so the prototypes of each cluster are meaningful and make possible to interpret the result.

Furthermore, we enhanced our profiling of the cryptocurrency market with association tests that validate the potential relationship between the clustering results and other descriptive features of the cryptocurrencies (technological attributes, financial ratios and age). These tests make possible to determine whether some features are related with a particular financial performance detected by the clustering algorithms.

Our analysis confirmed that there is an underlying structure of the data. Each one of the clustering algorithms revealed different aspects of the cryptocurrency market. Furthermore, the combination of the different the clustering results proved valid to detect the main trends in the cryptocurrency market, but also particular behaviours beyond these trends.

Finally, the association tests served to better describe the resulting clusters by adding the significant relationships found with the financial ratios, technological attributes and age of the cryptocurrencies.

In summary, we believe that the methodology used provides a descriptive tool supported by modern clustering techniques that may be useful for investors that need to understand the cryptocurrency market, as it reduces the dimensionality of the data set and identify the main trends in a descriptive manner. A straightforward extension of this research would include considering a longer time-frame for the data set just to investigate the dynamic evolution of the market structure and clusters beyond 2018. For further investigations, the associations of some of the key financial ratios and cluster associations could play an important role enhancing the performance of the algorithms for the asset selection and diversification of portfolios (Liu (2019), Platanakis et al. (2018), Brauneis and Mestel (2019)) or improving the forecasting performance (Mallikarjuna and Rao (2019)) to tackle the difficulty of a non-homogenous market.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

The initial idea was conceived by JA. The experiments were designed by LL. The searching on the data bases, statistical analysis and software design as performed by LL. The work was drafted by LL and revised critically by JA. All authors read and approved the final manuscript.

Author details

¹Complutense University, Faculty of Statistical Studies, Madrid, Spain. ²Complutense University, Faculty of Computation Sciences, Madrid, Spain.

References

- Acharya, J.G.A.: Cluster ensembles. *WIREs Data Mining and Knowledge discovery* **1**(4), 305–315 (2011)
- Aggarwal, C., C., Reddy, K., C.: *Data Clustering: Algorithms and Applications*, 1st edn. Chapman & Hall/CRC, ??? (2013)
- Aghabozorgi, S., Teh, Y.W.: Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications* **41**(4, Part 1), 1301–1314 (2014). doi:[10.1016/j.eswa.2013.08.028](https://doi.org/10.1016/j.eswa.2013.08.028)
- Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y.: Time-series clustering—a decade review. *Information Systems* **53**, 16–38 (2015)
- Agresti, A.: *An Introduction to Categorical Data Analysis*. John Wiley & Sons, ??? (2018)
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. *Pattern Recognition* **46**, 243–256 (2013)
- Arroyo, J., Maté, C.: Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting* **25**(1), 192–207 (2009). doi:[10.1016/j.ijforecast.2008.07.003](https://doi.org/10.1016/j.ijforecast.2008.07.003)
- Arroyo, J., González-Rivera, G., Maté, C., San Roque, A.M.: Smoothing methods for histogram-valued time series: an application to value-at-risk. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **4**(2), 216–228 (2011). doi:[10.1002/sam.10114](https://doi.org/10.1002/sam.10114)
- Bacon, C.R.: *Practical Portfolio Performance Measurement and Attribution*. The Wiley Finance Series. John Wiley & Sons, ??? (2008)
- Baek, C., Elbeck, M.: Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters* **22**(1), 30–34 (2015)
- Banerjee, A., Dave, R.N.: Validating clusters using the hopkins statistic. 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542) **1**, 149–153 (2004)
- Begum, N., Ulanova, L., Wang, J., Keogh, E.: Accelerating dynamic time warping clustering with a novel admissible pruning strategy. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15, pp. 49–58. Association for Computing Machinery, New York, NY, USA (2015). doi:[10.1145/2783258.2783286](https://doi.org/10.1145/2783258.2783286). <https://doi.org/10.1145/2783258.2783286>
- Begusic, S., Kostanjcar, Z., Eugene Stanley, H., Podobnik, B.: Scaling properties of extreme price fluctuations in bitcoin markets. *Physica A: Statistical Mechanics and its Applications* **510**, 400–406 (2018). doi:[10.1016/j.physa.2018.06.131](https://doi.org/10.1016/j.physa.2018.06.131)
- Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: *KDD Workshop*, vol. 10, pp. 359–370 (1994). Seattle, WA
- Bojanowski, M., Edwards, R.: *alluvial: R Package for Creating Alluvial Diagrams*. (2016). R package version: 0.1-2. <https://github.com/mbojan/alluvial>
- Bonanno, G., Caldarelli, G., Lillo, F., S., M., Vandewalle, N., Mantegna, R.N.: Networks of equities in financial markets. *The European Physical Journal B - Condensed Matter* **38**(2), 363–371 (2004). doi:[10.1140/epjb/e2004-00129-6](https://doi.org/10.1140/epjb/e2004-00129-6)

- Brauneis, A., Mestel, R.: Cryptocurrency-portfolios in a mean-variance framework. *Finance Research Letters* **28**, 259–264 (2019)
- Brida, J., Risso, W.: Dynamics and structure of the 30 largest north american companies. *Society for Computational Economics* **35**(1), 85–99 (2009)
- Burniske, C., Tatar, J.: *Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyond*. McGraw-Hill Education, ??? (2017). <https://books.google.es/books?id=-5AtDwAAQBAJ>
- Chan, S., Chu, J., Nadarajah, S., Osterrieder, J.: A statistical analysis of cryptocurrencies. *Journal of Risk and Financial Management* **10**(2), 12 (2017)
- Chaudhuri, T.D., Ghosh, I.: Using clustering method to understand Indian stock market volatility. *Communications on Applied Electronics* **2**(6), 35–44 (2015)
- Corbet, S., Meegan, A., Larkin, C.J., Lucey, B., Yarovaya, L.: Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economic Letters* **165**, 28–34 (2018)
- Corbet, S., Lucey, B., Urquhart, A., Yarovaya, L.: Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis* **62**, 182–199 (2019)
- Drozd, S., Minati, L., Oswiecimka, P., Stanuszek, M., Watorek, M.: Signatures of the crypto-currency market decoupling from the forex. *Future Internet* **11**(7), 154 (2019). doi:[10.3390/fi11070154](https://doi.org/10.3390/fi11070154)
- D'Urso, P., De Giovanni, L., Massari, R.: Garch-based robust clustering of time series. *Fuzzy Sets Syst.* **305**(C), 1–28 (2016). doi:[10.1016/j.fss.2016.01.010](https://doi.org/10.1016/j.fss.2016.01.010)
- D'Urso, P., Giovanni, L.D., Massari, R., D'Ecclesia, R.L., Maharaj, E.A.: Cepstral-based clustering of financial time series. *Expert Systems with Applications* **161**, 113705 (2020). doi:[10.1016/j.eswa.2020.113705](https://doi.org/10.1016/j.eswa.2020.113705)
- Fisher, R.A.: On the interpretation of the χ^2 from contingency tables and the calculation of the p. *Royal Statistical Society* **85**(1), 87–94 (1922)
- Gillespie, C.S.: Fitting heavy tailed distributions: The powerLaw package. *Journal of Statistical Software* **64**(2), 1–16 (2015)
- González-Rivera, G., Arroyo, J.: Time series modeling of histogram-valued data: The daily histogram time series of s&p500 intradaily returns. *International Journal of Forecasting* **28**(1), 20–33 (2012). doi:[10.1016/j.ijforecast.2011.02.007](https://doi.org/10.1016/j.ijforecast.2011.02.007)
- Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**(1), 100–108 (1979)
- Henning, C., Meila, M., Murtagh, F., Rocci, R.: *Handbook of Cluster Analysis*. CRC Press, ??? (2016)
- Hornik, K.: A CLUE for CLUster Ensembles. *Journal of Statistical Software* **14**(12) (2005). doi:[10.18637/jss.v014.i12](https://doi.org/10.18637/jss.v014.i12)
- Hornik, K.: Clue: Cluster Ensembles. (2019). R package version 0.3-57. <https://CRAN.R-project.org/package=clue>
- Hu, A.S., Parlour, C.A., Rajan, U.: Cryptocurrencies: Stylized facts on a new investible instrument. *Financial Management* **48**(4), 1049–1068 (2019)
- Irpino, A.: Histogram-value data analysis. R library **1**, (2019)
- Irpino, A., Verde, R.: Dynamic clustering of histograms using wasserstein metric. In: *COMPSTAT 2006, Proceedings in Computational Statistics*, pp. 869–876. Physica-Verlag, Heidelberg (2006)
- Irpino, A.: HistDAWass Package: An R Tool for Histograms-values Data. (2016). R package version 1.0.4. <https://cran.r-project.org/package=HistDAWass>
- Irpino, A., Verde, R.: Basic statistics for distributional symbolic variables: a new metric-based approach. *Advances in Data Analysis and Classification* **9**, 143–175 (2015)
- Irpino, A., Verde, R., De Carvalho, F.d.A.T.: Dynamic clustering of histogram data based on adaptive squared Wasserstein distances. *Expert Systems with Applications* **41**(7), 3351–3366 (2014). doi:[10.1016/j.eswa.2013.12.001](https://doi.org/10.1016/j.eswa.2013.12.001)
- Kern, M., Lex, A., Gehlenborg, N., Johnson, C.R.: Interactive visual exploration and refinement of cluster assignments. *BMC bioinformatics* **18**(1), 1–13 (2017)
- Lemire, D.: Faster retrieval with a two-pass dynamic-time-warping lower bound. *CoRR* **abs/0811.3301** (2008). doi:[10.1145/1471714](https://doi.org/10.1145/1471714)
- Letra, I.J.S.: What drives cryptocurrency value? a volatility and predictability analysis. PhD thesis, Instituto Superior de Economia e Gestão (2016)
- Liao, S.-H.: Mining stock category association and cluster on Taiwan stock market. *Expert Systems with Applications* **35**, 19–29 (2007)
- Liao, S.-H., Chou, S.-Y.: Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio. *Expert Systems with Applications* **40**(5), 1542–1554 (2013). doi:[10.1016/j.eswa.2012.08.075](https://doi.org/10.1016/j.eswa.2012.08.075)
- Liao, T.W.: Clustering of time series data-a survey. *The journal of the pattern recognition society* **1**(38), 1857–1874 (2005)
- Liu, W.: Portfolio diversification across cryptocurrencies. *Finance Research Letters* **29**, 200–205 (2019)
- L'Yi, S., Ko, B., Shin, D., Cho, Y.-J., Lee, J., Kim, B., Seo, J.: XCluSim: a visual analytics tool for interactively comparing multiple clustering results of bioinformatics data. *BMC bioinformatics* **16**(S11), 5 (2015)
- MacQueen, J.: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**(-), 281–297 (1967)
- Mallikarjuna, M., Rao, R.P.: Evaluation of forecasting methods from selected stock market returns. *Financial Innovation* **5**(1), 1–16 (2019). doi:[10.1186/s40854-019-0157-x](https://doi.org/10.1186/s40854-019-0157-x)
- Mantegna, R.: Hierarchical structure in financial markets. *European Physical Journal B* **11**(1), 193–197 (1999)
- Marti, G., Nielsen, F., Bińkowski, M., Donnat, P.: A review of two decades of correlations, hierarchies, networks and clustering in financial markets. *Papers* 1703.00485, arXiv.org (March 2017). <https://arxiv.org/abs/1703.00485>
- Mehta, C.R., Patel, N.R.: A network algorithm for performing fisher exact test in rxc contingency table. *Journal of the American Statistical Association* **78**(382), 427–434 (1983)

- Mehta, C.R., Patel, N.R.: Exact testsTM. SPSS exact tests **7**, 12 (1996)
- Meyer, D., Zeileis, A., Hornik, K.: The strucplot framework: Visualizing multi-way contingency tables with vcd. *Journal of Statistical Software* **17**(3), 1–48 (2006)
- Meyer, D., Zeileis, A., Hornik, K.: Vcd: Visualizing Categorical Data. (2020). R package version 1.4-7
- Mizuno, T., Takayasu, H., Takayasu, M.: Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications* **364**, 336–342 (2006). doi:[10.1016/j.physa.2005.08.079](https://doi.org/10.1016/j.physa.2005.08.079)
- Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2009). <http://www.bitcoin.org/bitcoin.pdf>
- Nanda, S.R., Mahanty, B., Tiwari, M.K.: Clustering Indian stock market data for portfolio management. *Expert System with Applications* **37**, 8793–8798 (2010)
- Newman, M.: Power laws, pareto distributions and zipf's law. *Contemporary Physics* **46**(5), 323–351 (2005). doi:[10.1080/00107510500052444](https://doi.org/10.1080/00107510500052444). <http://www.tandfonline.com/doi/pdf/10.1080/00107510500052444>
- Noirhomme-Fraiture, M., Brito, P.: Far beyond the classical data models: symbolic data analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **4**(2), 157–170 (2011). doi:[10.1002/sam.10112](https://doi.org/10.1002/sam.10112)
- Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., Kanto, A.: Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E* **68**(5) (2003). doi:[10.1103/physreve.68.056110](https://doi.org/10.1103/physreve.68.056110)
- Pele, D., Wesselhöft, N., Härdle, W., Kolosiat, M., Yannis, Y.: A Statistical Classification of Cryptocurrencies (2020). <https://ssrn.com/abstract=3548462>
- Peterson, B.G., Carl, P., Boudt, K., Bennet, R., Ulrich, J., Zivot, E., Lestel, M., Balkissoon, K., Wuertz, D.: PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. (2018). R package version 1.5.2. <https://cran.r-project.org/package=PerformanceAnalytics>
- Pierpaolo, D., Cappelli, C., Di Lallo, D., Massari, R.: Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications* **392**(9), 2114–2129 (2013)
- Platanakis, E., Sutcliffe, C., Urquhart, A.: Optimal vs naïve diversification in cryptocurrencies. *Economics Letters* **171**, 93–96 (2018)
- R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2013). R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rani, S., Sikka, G.: Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications* **52**(15) (2012)
- Rodríguez, A., Laio, A.: Clustering by fast search and find of density peaks. *Science* **344**(6191), 1492–1496 (2014). doi:[10.1126/science.1242072](https://doi.org/10.1126/science.1242072). <https://science.sciencemag.org/content/344/6191/1492.full.pdf>
- Rosvall, M., Bergstrom, C.T.: Mapping change in large networks. *PLoS ONE* **5** (2010)
- Sarda-Espinosa, A.: Dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance. (2019). R package version 5.5.6. <https://CRAN.R-project.org/package=dtwclust>
- Sardá-Espinosa, A.: Time-series clustering in R using the dtwclust package. *The R Journal* **11**(1), 22–43 (2019)
- Sigaki, H.Y.D., Perc, M., Ribeiro, H.V.: Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market. In: *Scientific Reports* (2019)
- Song, J.Y., Chang, W., Song, J.W.: Cluster analysis on the structure of the cryptocurrency market via Bitcoin-Ethereum filtering. *Physica A-Statistical Mechanics and its Applications* **527** (2019)
- Stosic, D., Stosic, D., Luderer, T.B., Stosic, T.: Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications* **507**, 499–509 (2018)
- Watorek, M., Drozd, S., Kwapien, J., Minati, L., Oswiecimka, P., Stanuszek, M.: Multiscale characteristics of the emerging global cryptocurrency market. *Physics Reports* (2020). doi:[10.1016/j.physrep.2020.10.005](https://doi.org/10.1016/j.physrep.2020.10.005)
- Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., et al.: Top 10 algorithms in data mining. *Knowledge and information systems* **14**(1), 1–37 (2008)
- Yates, F.: Tests of significance for 2 × 2 contingency tables. *Royal Statistical Society* **147**(3), 426–463 (1984)
- Yermack, D.: Is bitcoin a real currency? an economic appraisal. Working Paper 19747, National Bureau of Economic Research (December 2013). doi:[10.3386/w19747](https://doi.org/10.3386/w19747). <http://www.nber.org/papers/w19747>
- Zeileis, A., Meyer, D., Hornik, K.: Residual-based shadings for visualizing (conditional) independence. *Journal of Computational and Graphical Statistics* **16**(3), 507–525 (2007)
- Zhang, W., Wang, P., Li, X., Shen, D.: Some stylized facts of the cryptocurrency market. *Applied Economics* **50**(55), 5950–5965 (2018)
- Zieba, Damian, Kokoszczyski, Ryszard, Sledziwska, Katarzyna: Shock transmission in the cryptocurrency market. is bitcoin the most influential? *International Review of Financial Analysis* **64**, 102–125 (2019)
- @settingslabel, options="nameyear"

Figures

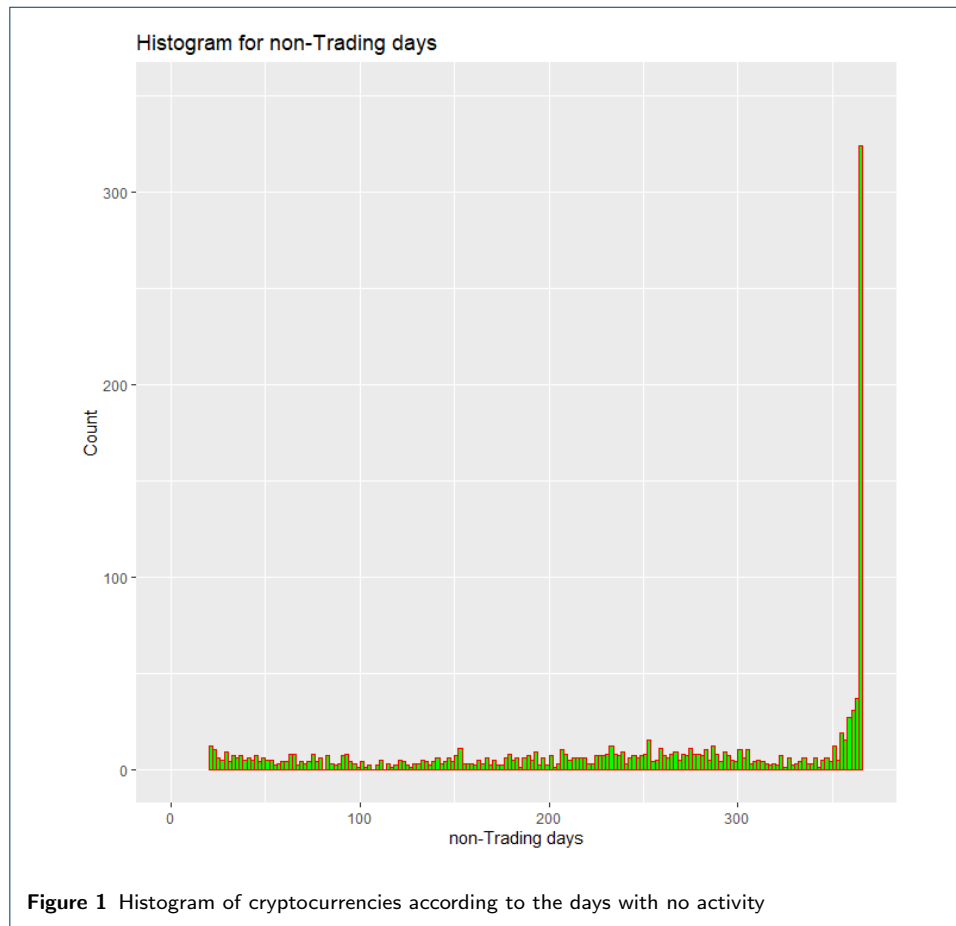
Tables

Additional Files

Additional file 1 — Sample additional file title

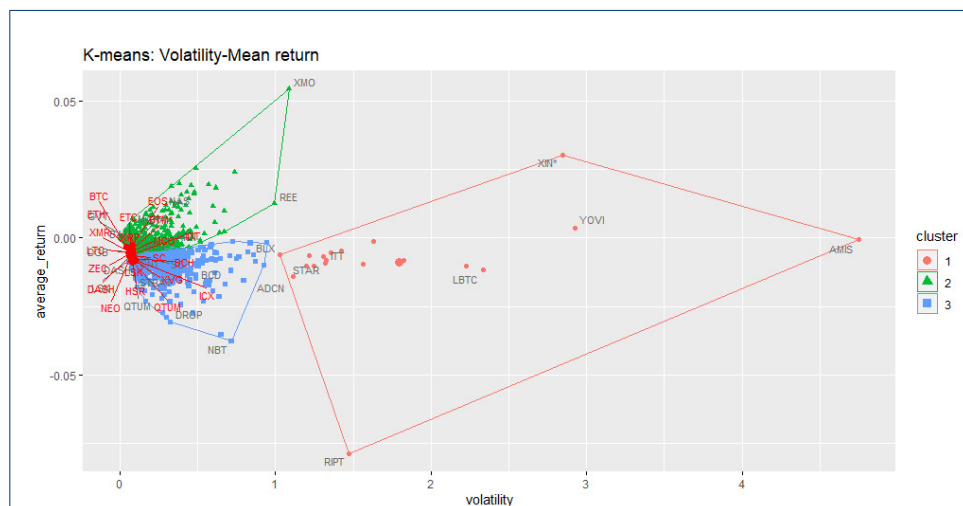
Additional file 2 — Sample additional file title

Additional file descriptions text.

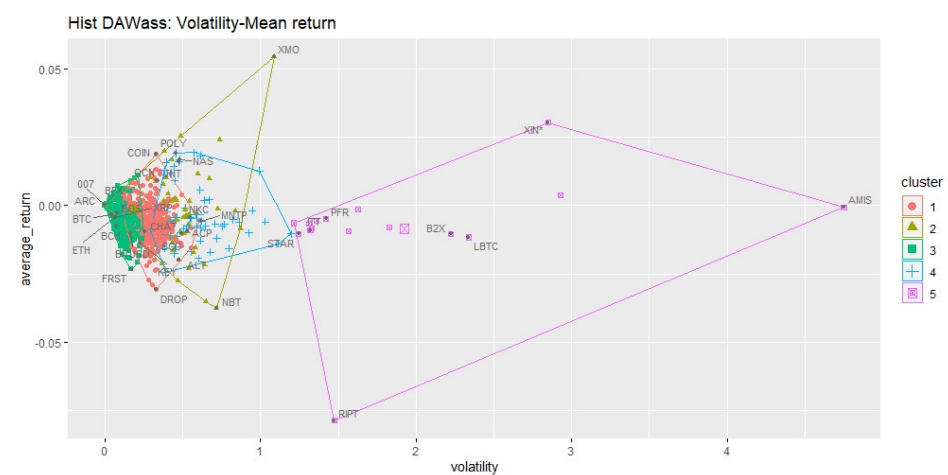


Variable	# Levels	Values
Algorithm	73	Encryption algorithm (SHA256, Ethash, X13, X11,...)
ProofType	39	Consensus algorithm (PoW, PoW/PoS,DPoS..)
Volume	5	Percentiles of the volume negotiated. Namely, P_{70} for volume values lower than the P_{70} percentile, P_{80} for values higher than the P_{70} and lower than the P_{90} , and similarly P_{90} , P_{99} and P_{100} .
MkCap	5	Percentiles of the market capitalization. Namely, P_{70} for market cap values lower than the P_{70} percentile, P_{80} for values higher than the P_{70} and lower than the P_{90} , and similarly P_{90} , P_{99} and P_{100} .
Beta	6	Beta values divided into the following categories: <i>NegBeta</i> for beta values lower than -0.01 <i>CashLike</i> if beta is to equal or higher than -0.01 and lower than 0.01 <i>LowVol</i> if beta is equal to or higher than 0.01 and lower than 0.95 <i>Indexlike</i> if beta is equal to or higher than 0.95 and lower than 1.05 <i>HighVol</i> if beta is equal to or higher than 1.05 and lower than 100 <i>Extreme</i> if beta is higher than 100
Sharpe	6	Sharpe ratio divided into the following categories: <i>SRF</i> (Small Risk-free) for negative values <i>ERP</i> (Excess return positive) for positive values lower than 0.5 <i>ACC</i> (Acceptable) for values equal to or higher than 0.5 and lower than 1.0 <i>GOOD</i> for values equal to or higher than 1.0
Age	7	Deciles of the age variable (time on the market). We use the same partition than in the M2 ratio.

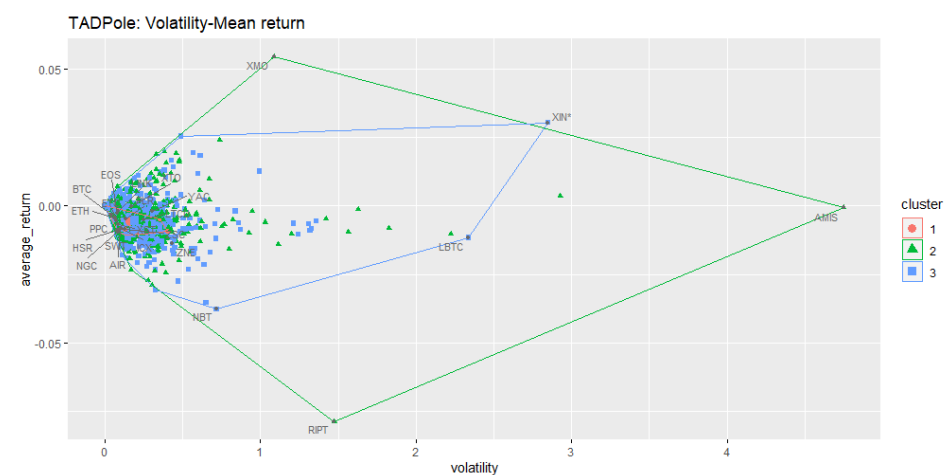
Table 1 Categorical variables used on the association tests and its values (*NA* was used for not available values)



(a) K-means clustering for ordinary variables with the position of highest market cap in red colours



(b) Histogram DAWass clustering for ordinary variables



(c) TADPole clustering for ordinary variables

Figure 2 Volatility-Average return (two-moments) plane in ordinary values with the vertex names and the more representative cryptocurrencies in terms of market cap for the different clustering techniques

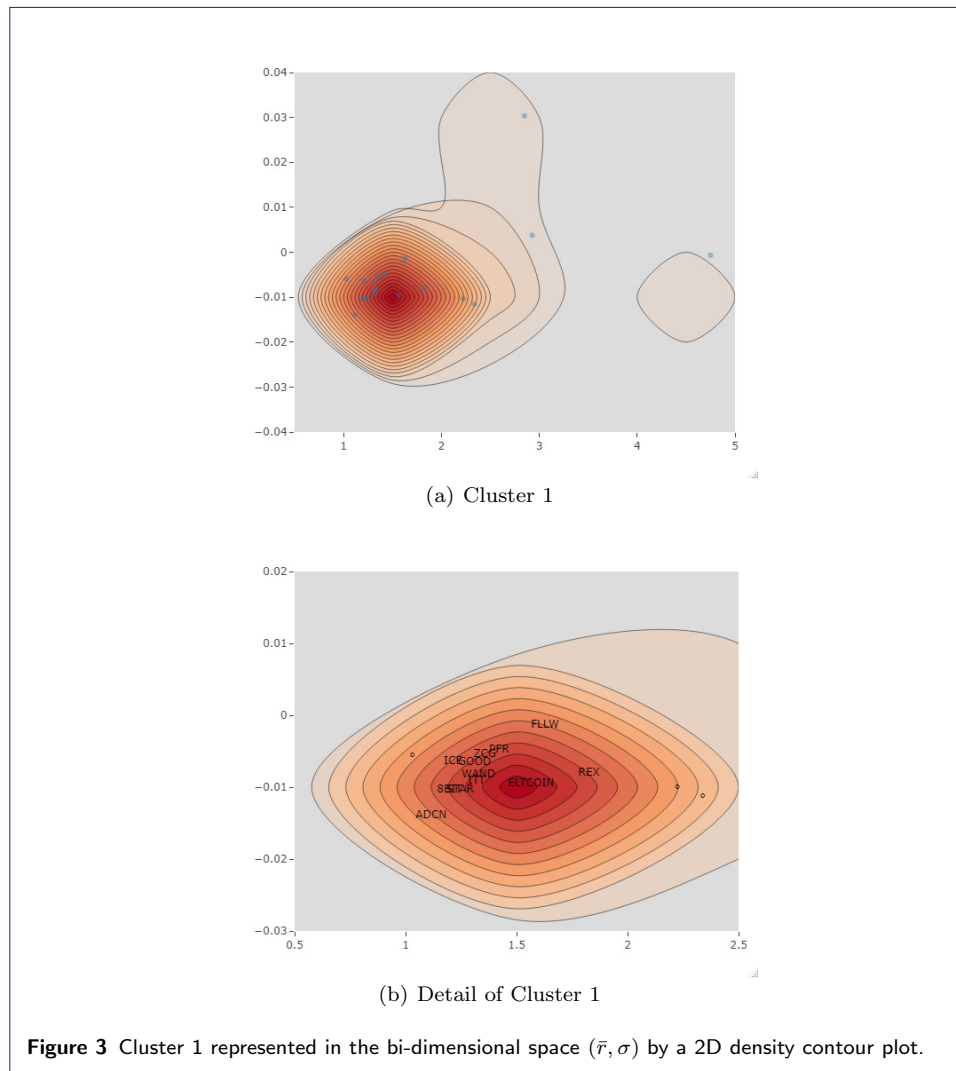


Figure 3 Cluster 1 represented in the bi-dimensional space (\bar{r}, σ) by a 2D density contour plot.

	K-means			HistDAWass			TADPole		
	Card.	Mean	Std.Dev.	Card.	Mean	Std.Dev.	Card.	Mean	Std.Dev.
Clus. 1	19	-0.008	1.795	496	-0.134	0.337	22	-0.001	0.080
Clus. 2	903	-0.002	0.130	147	-0.503	0.378	843	0.026	0.046
Clus. 3	801	-0.009	0.229	1007	-0.011	0.108	858	-0.028	0.047
Clus. 4				57	-0.044	0.867			
Clus. 5				16	-0.095	3.123			

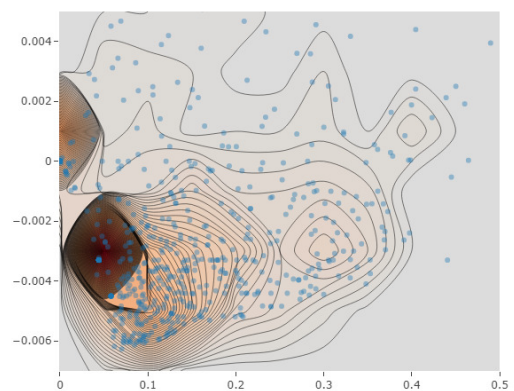
Table 2 Cluster cardinality, mean value and standard deviation of the centroid or prototypes for the clustering methods. For Hist DAWass and TADPole we compute the mean and standard deviation of the prototypes.

	Mean	Std. Dev.	Coef.Var.	Skew.	Kurt.	Med.	Min.	Max.	Var.Wass.
Clus. 1	-0.13	0.34	-2.51	0.82	13.43	-0.16	-2.24	2.36	0.025
Clus. 2	-0.50	0.38	-0.75	0.56	9.33	-0.51	-2.69	2.18	0.079
Clus. 3	-0.01	0.11	-10.06	0.28	7.10	-0.01	-0.55	0.62	0.005
Clus. 4	-0.04	0.87	-19.97	0.54	11.95	-0.08	-5.44	6.67	0.128
Clus. 5	-0.09	3.12	-32.90	0.05	5.66	-0.17	-17.56	17.56	1.116

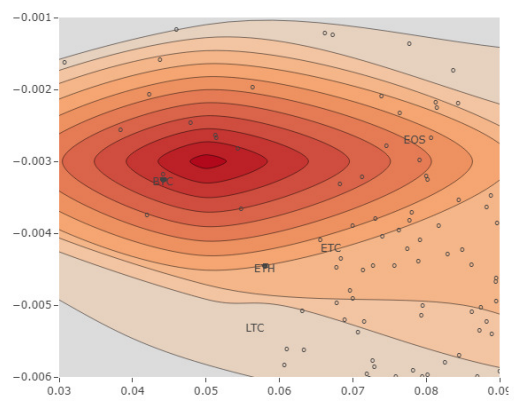
Table 3 Descriptive statistics for the prototypes of the Hist DAWass clustering.

Cluster	Mean Dist.	Std. Dev.	Coef. Var.
1	4.31	3.04	0.71
2	4.60	3.29	0.72
3	4.85	3.53	0.73

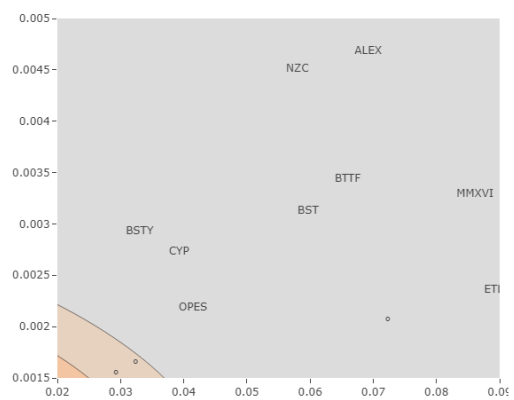
Table 4 Variability of TADPole clusters with the mean distance (Mean Dist.) to the centroid, standard deviation (Std. Dev.) and coefficient of variation (Coef. Var.).



(a) Cluster 2

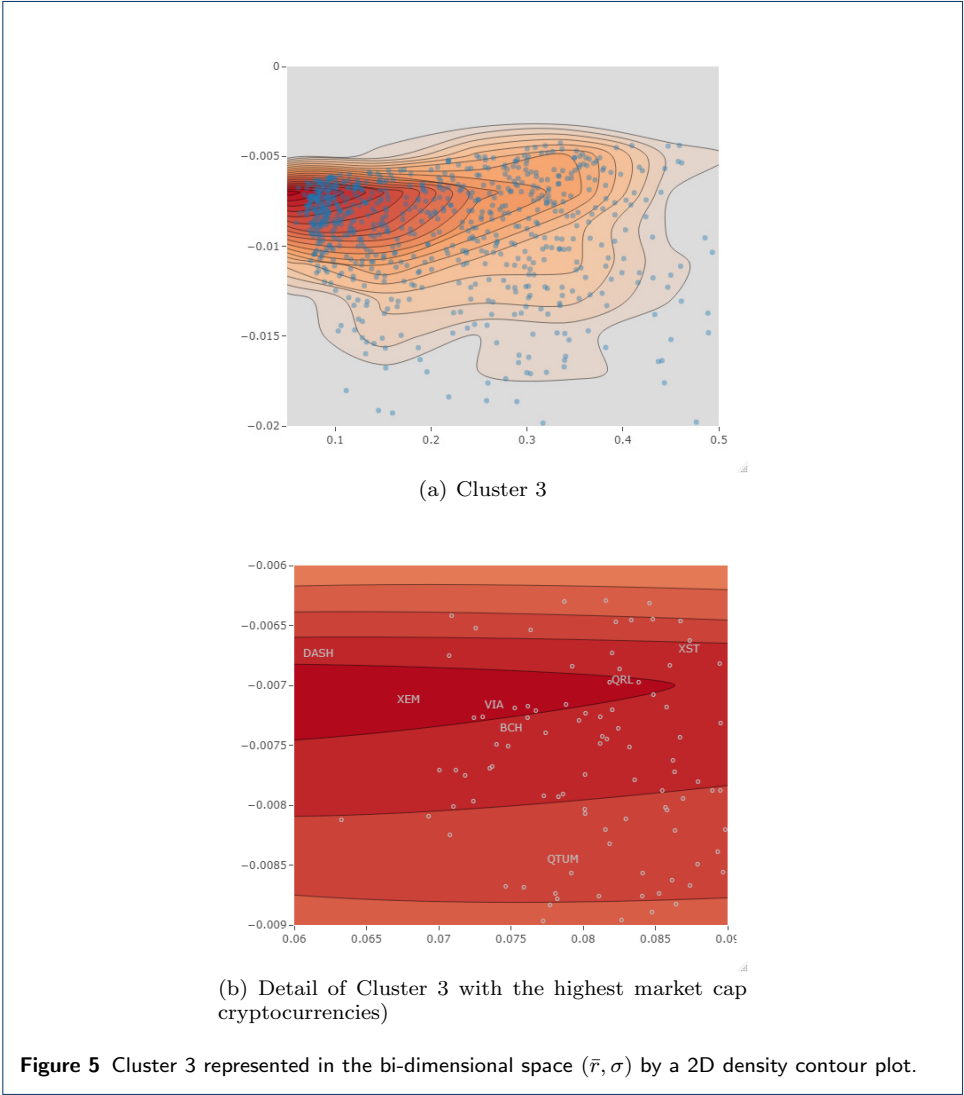


(b) Detail of Cluster 2 with the highest market cap cryptocurrencies



(c) Detail of cluster 2 for the high returns and low volatility cryptocurrencies

Figure 4 Cluster 2 represented in the bi-dimensional space (\bar{r}, σ) by a 2D density contour plot.



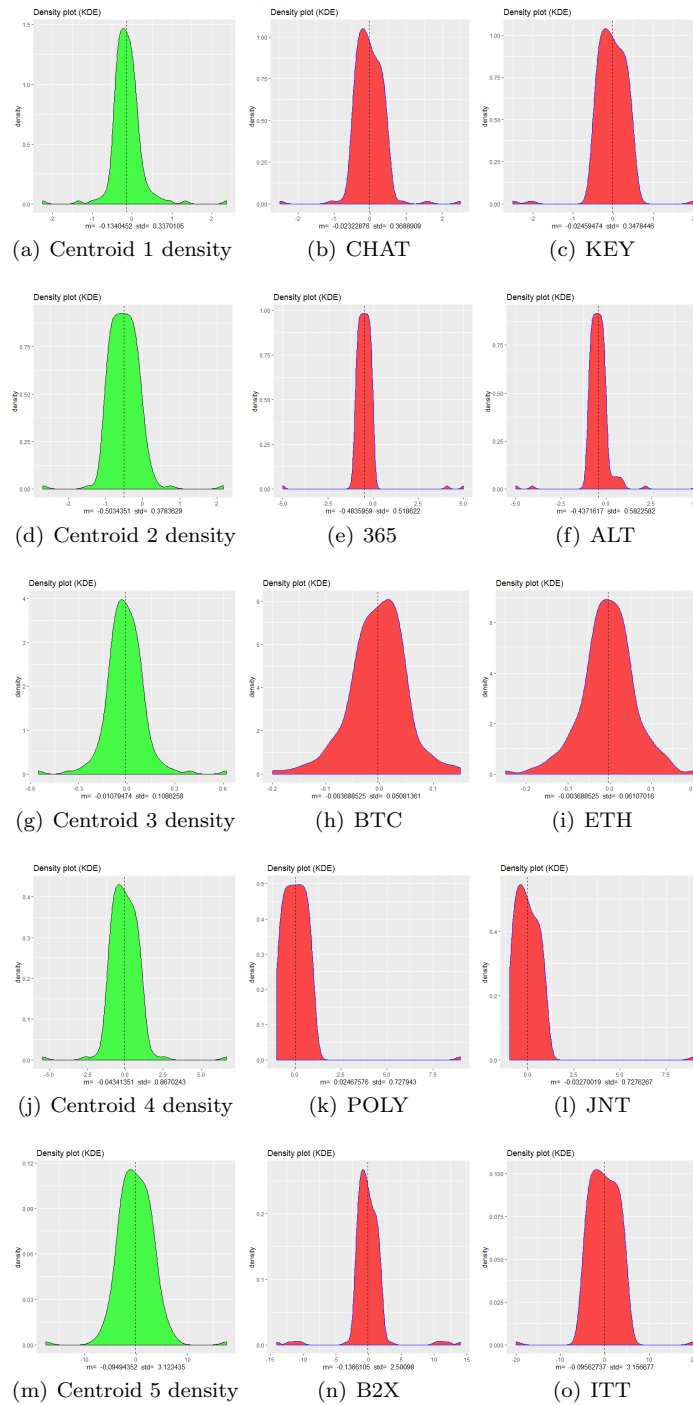
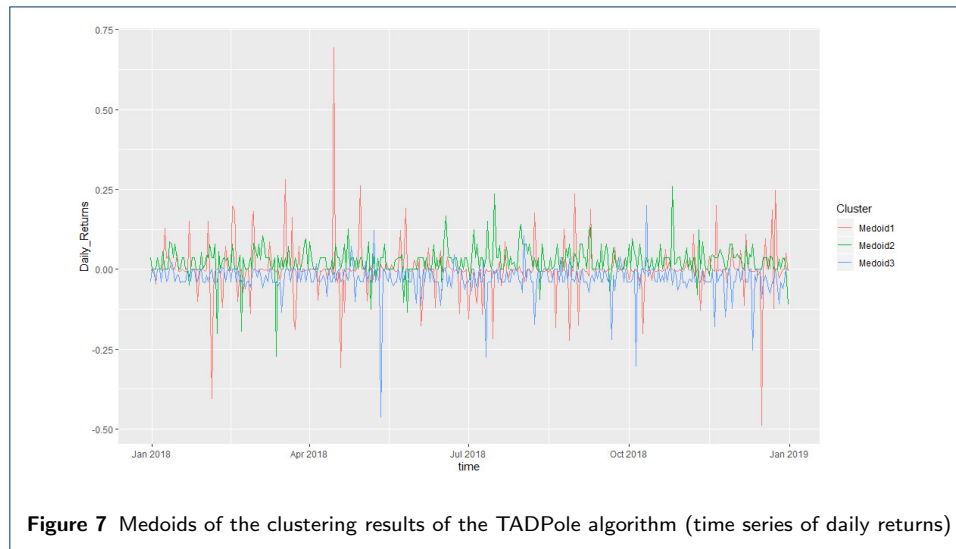
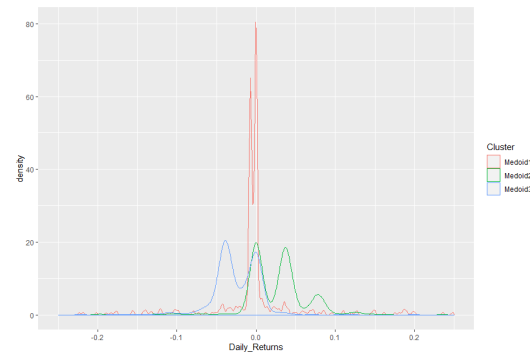


Figure 6 Density plot for prototypes (first column), and some representative cryptocurrencies of each cluster in terms of market capitalization (2nd and 3rd columns)

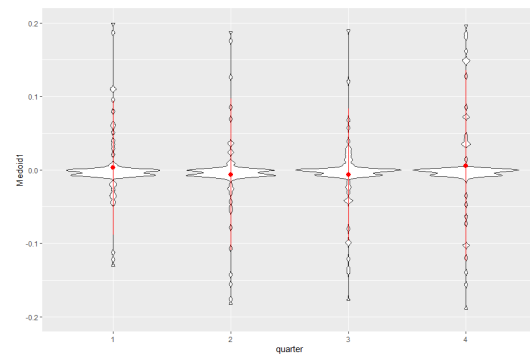


Intersection	Kmeans	HistDAWass	TADPole	Combi	N
1	2	3	3	1	295
2	2	3	2	2	294
3	3	3	3	3	208
4	3	3	2	4	196
5	3	1	3	5	166
6	3	1	2	6	148
7	2	1	2	7	97
8	2	1	3	8	78
9	2	2	3	9	57
10	2	2	2	10	54
11	3	2	3	11	20
12	3	4	2	12	18
13	3	4	3	13	18
14	3	2	2	14	15
15	2	4	2	15	10
16	2	4	3	16	8
17	1	5	2	17	8
18	1	5	3	18	8
19	2	3	1	19	7
20	3	3	1	20	7
21	3	1	1	21	5
22	1	4	2	22	3
23	2	1	1	23	2
24	2	2	1	24	1

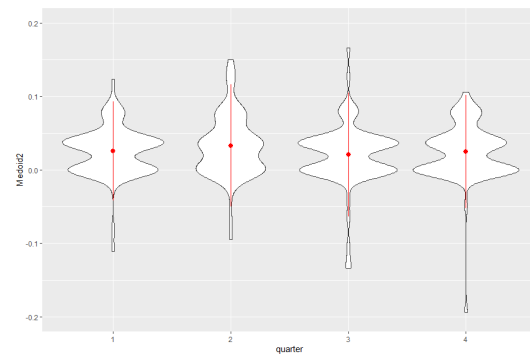
Table 5 Intersection of clusters across the different clustering algorithms, each column represent the cluster number. Intersections are sorted in inverse cardinality (N) order.



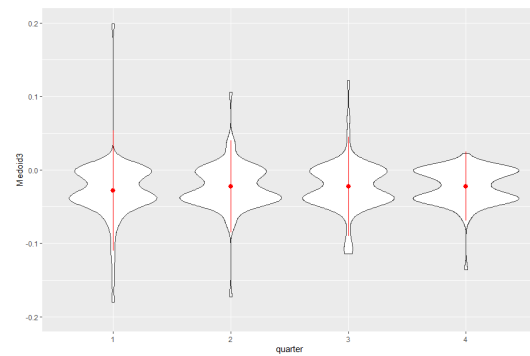
(a) Density plot of the three medoids



(b) Quarterly distribution of the Cluster 1 medoid (LINK)

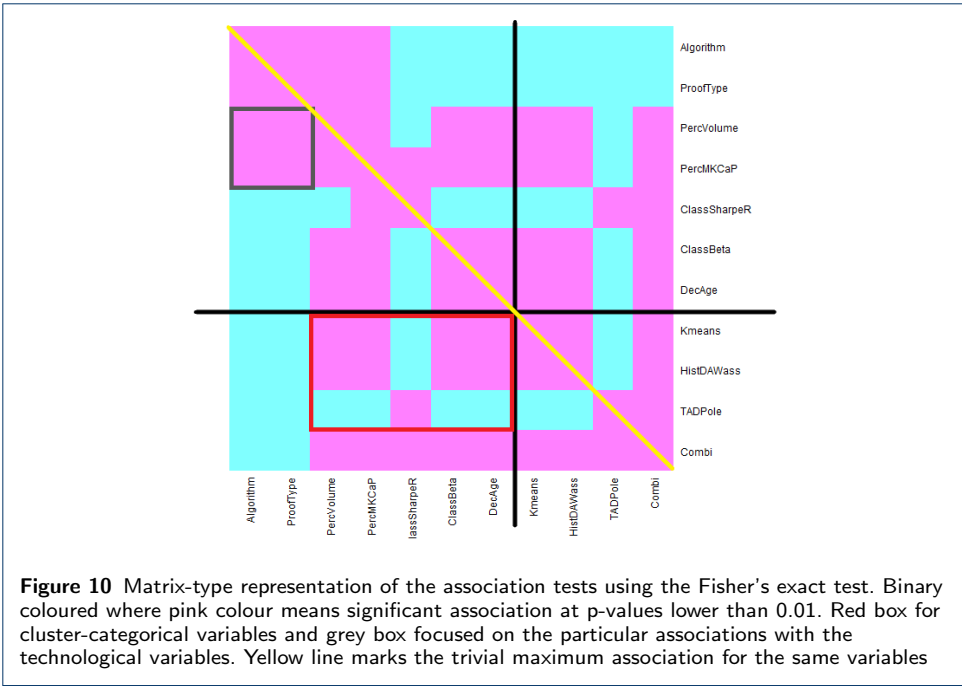
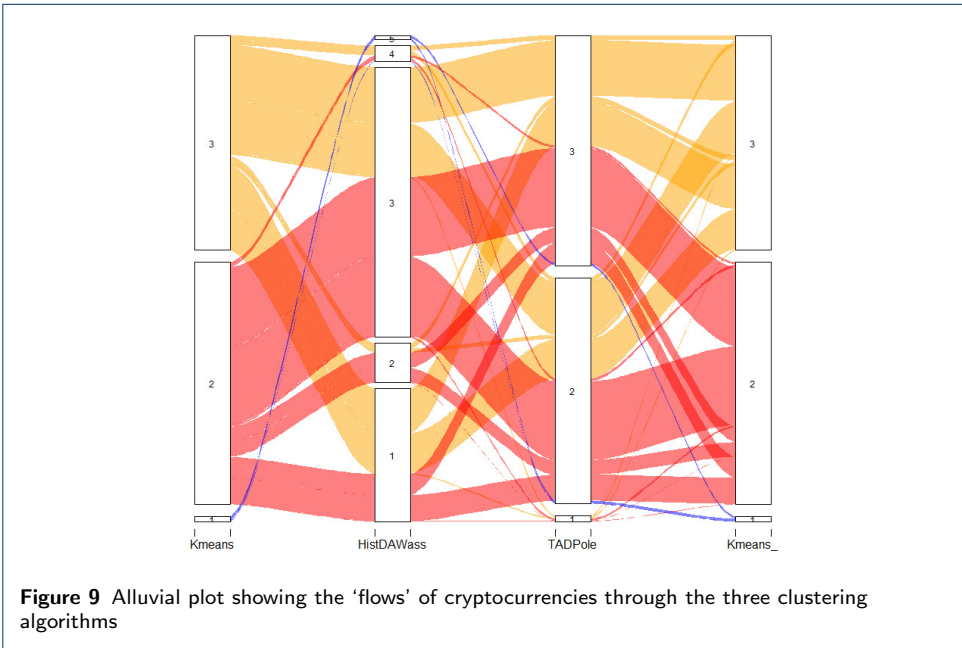


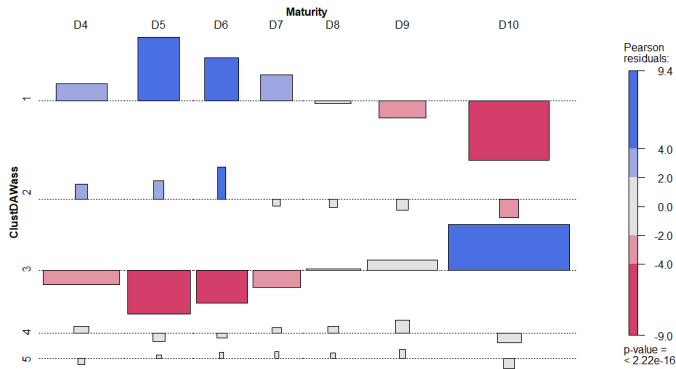
(c) Quarterly distribution of the Cluster 2 medoid (XTO)



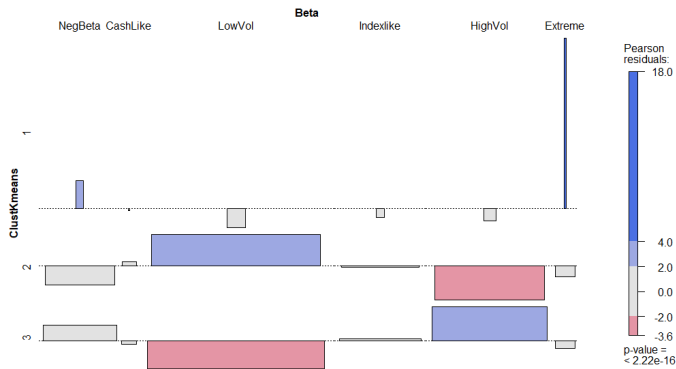
(d) Quarterly distribution of the Cluster 1 medoid (ZNE)

Figure 8 Density plot in Fig.a) and quarterly distribution of the TADPole medoids represented by violin graphs in Fig. b), c) and d) with the mean and standard deviation.

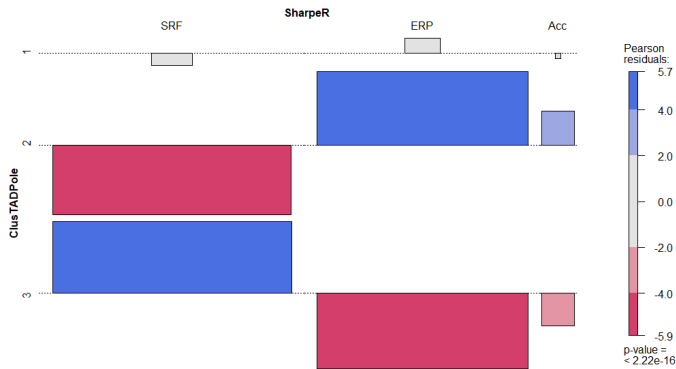




(a) Maturity VS Hist DAWass clustering



(b) Beta VS K-means



(c) Sharpe ratio VS TADPole

Figure 11 Pearson's residual representation for combination of different clustering techniques and categoric variables

Technique	Cluster/Intersection	P70	P80	P90	P99	P100
PercVolume VS K-means	1	-0.13	2.43	-1.05	-0.97	-0.43
	2	8.93	-4.73	-4.20	-5.57	2.61
	3	-8.90	4.36	4.35	5.71	-2.54
PercVolume VS Hist DAWass	1	12.25	-4.33	-4.82	-7.26	-3.67
	2	3.72	-1.71	-1.78	-1.66	-0.74
	3	-12.01	3.09	5.09	7.85	3.71
	4	-2.02	3.36	0.44	-0.95	0.22
	5	-0.48	2.78	-0.97	-0.90	-0.40
PercMKCaP VS K-means	1	1.15	0.28	-0.96	-0.91	-0.37
	2	3.91	-5.90	-1.77	0.20	3.57
	3	-4.08	5.85	1.92	-0.06	-3.51
PercMKCaP VS Hist DAWass	1	8.07	-1.70	-4.98	-4.64	-2.21
	2	2.36	-0.87	-1.63	-0.81	-0.63
	3	-8.37	1.88	4.89	4.84	2.59
	4	-0.44	-0.24	1.37	-0.16	-0.78
	5	0.94	0.44	-0.89	-0.84	-0.34
PercVolume VS Combi	1	3.66	-2.20	-2.62	-2.52	4.02
	2	5.98	-2.52	-2.51	-3.49	-0.41
	3	-10.25	2.57	6.21	6.21	-0.26
	4	-12.42	6.39	4.53	7.35	-0.21
	5	7.16	-3.21	-2.24	-3.95	-2.14
	6	7.30	-1.25	-3.96	-4.39	-1.97
PercMKCap VS Combi	1	2.27	-3.02	-2.45	0.86	2.97
	2	2.17	-3.34	0.40	-0.90	1.31
	3	-6.69	3.98	4.80	1.89	-1.54
	4	-6.33	3.63	2.72	3.31	-0.33
	5	4.88	-0.99	-3.38	-2.22	-1.72
	6	4.87	0.21	-2.94	-3.96	-1.58

Table 6 Volume and Market cap - Standardized Person's residuals

Technique	Cluster/Intersection	SRF	ERP	Acc
TADPole	1	-1.43	1.52	-0.44
	2	-11.32	10.60	3.69
	3	11.65	-10.95	-3.58
Combi	1	5.83	-5.49	-1.74
	2	-6.02	5.20	4.13
	3	3.92	-3.63	-1.53
	4	-4.67	4.67	0.11
	5	3.93	-3.68	-1.24
	6	-3.00	3.04	-0.15

Table 7 Sharpe ratio - Standardized Person's residuals

Technique	Cluster/Intersection	NegBeta	CashLike	LowVol	Indexlike	HighVol	Extreme
K-means	1	3.05	-0.17	-2.92	-0.97	-1.45	17.79
	2	-2.87	0.57	6.70	-0.18	-5.58	-1.51
	3	2.41	-0.54	-6.26	0.32	5.79	-1.13
Hist DAWass	1	12.09	1.57	-0.32	-5.84	-3.37	-1.90
	2	3.97	-0.28	-2.24	-0.94	0.74	-0.38
	3	-15.24	-1.29	3.10	6.66	2.83	-4.29
	4	6.93	-0.36	-5.47	-2.05	1.23	10.33
	5	2.02	-0.15	-2.70	-0.89	-1.34	19.24
Combi	1	-3.71	-0.90	4.50	-0.19	-2.92	-
	2	-4.07	0.49	5.62	0.57	-4.82	-
	3	-3.47	-0.79	-5.74	3.28	6.20	-
	4	-3.42	0.76	-5.41	1.99	6.62	-
	5	7.87	-0.65	1.73	-3.61	-3.53	-
	6	10.52	1.29	-1.69	-3.16	-1.61	-

Table 8 Beta - Standardized Person's residuals

Technique	Cluster/Intersection	D4	D5	D6	D7	D8	D9	D10
K-means	1	-1.14	0.32	0.76	2.48	0.63	1.18	-2.14
	2	4.74	-0.34	-1.75	0.91	1.33	-1.14	-2.79
	3	-4.56	0.29	1.63	-1.29	-1.43	0.96	3.11
Hist DAWass	1	3.27	11.88	7.91	4.70	-0.51	-3.35	-13.70
	2	2.41	2.85	4.97	-1.15	-1.32	-1.78	-3.64
	3	-3.99	-11.80	-8.80	-4.65	0.38	2.80	15.08
	4	1.08	-1.35	-0.83	0.83	1.06	2.05	-1.95
	5	-1.06	0.49	0.94	1.08	0.80	1.43	-1.98
Combi	1	1.70	-2.76	-1.40	0.19	0.24	0.36	0.51
	2	4.71	-1.10	-1.73	-0.15	1.93	-1.22	-2.13
	3	-5.47	-4.02	-3.05	-1.46	-1.82	3.07	7.20
	4	-5.61	-4.24	-2.99	-2.84	-0.54	1.55	8.37
	5	1.35	6.34	7.38	2.87	1.16	-2.07	-8.53
	6	3.87	8.38	3.62	2.16	-1.15	-2.38	-7.92

Table 9 *Maturity* - Standardized Person's residuals

	P70	P80	P90	P99	P100
DPoR	-1.20	-0.38	2.51	-0.37	-0.16
DPoS	-2.42	-0.45	-0.54	3.18	3.15
DPoS/LPoS	1.18	-0.54	-0.56	-0.52	-0.23
dPoW/PoW	-1.20	-0.38	-0.40	2.71	-0.16
LFT	-1.20	-0.38	-0.40	-0.37	6.23
LPoS	-1.20	-0.38	2.51	-0.37	-0.16
N/A	-9.62	2.12	5.00	6.76	0.69
PoA	0.83	-0.38	-0.40	-0.37	-0.16
PoB/PoS	0.83	-0.38	-0.40	-0.37	-0.16
POBh	-1.20	2.63	-0.40	-0.37	-0.16
PoC	-0.91	-0.66	0.99	1.14	-0.28
PoI	-1.20	-0.38	2.51	-0.37	-0.16
PoP	0.83	-0.38	-0.40	-0.37	-0.16
PoP/PoV/PoQ	0.83	-0.38	-0.40	-0.37	-0.16
PoPP	-1.20	-0.38	-0.40	2.71	-0.16
PoS	3.66	-1.65	-1.41	-1.74	-1.29
PoS/LPoS	-1.20	-0.38	-0.40	2.71	-0.16
PoS/PoB	0.83	-0.38	-0.40	-0.37	-0.16
PoS/PoW	0.83	-0.38	-0.40	-0.37	-0.16
PoS/PoW/PoT	-1.20	2.63	-0.40	-0.37	-0.16
PoSign	0.83	-0.38	-0.40	-0.37	-0.16
PoST	-1.20	-0.38	-0.40	2.71	-0.16
PoW	1.43	0.69	-0.73	-2.68	1.23
PoW and PoS	0.83	-0.38	-0.40	-0.37	-0.16
PoW/HiPoS	0.83	-0.38	-0.40	-0.37	-0.16
PoW/nPoS	-1.20	-0.38	-0.40	2.71	-0.16
PoW/PoM/PoSII	0.83	-0.38	-0.40	-0.37	-0.16
PoW/PoS	7.34	-1.51	-3.80	-4.65	-1.84
PoW/PoS	-0.26	1.59	-0.56	-0.52	-0.23
PoW/PoS	-1.20	-0.38	-0.40	2.71	-0.16
PoWT	-1.20	-0.38	2.51	-0.37	-0.16
Proof of Authority	-1.20	-0.38	2.51	-0.37	-0.16

Table 10 *Consensus algorithm - Volume* Standardized Person's residuals

	P70	P80	P90	P99	P100
DPoR	-1.36	2.83	-0.36	-0.34	-0.13
DPoS	-1.69	-0.31	-0.36	1.64	3.91
DPoS/LPoS	1.04	-0.50	-0.51	-0.49	-0.19
dPoW/PoW	-1.36	-0.35	-0.36	2.91	-0.13
LFT	-1.36	-0.35	-0.36	2.91	-0.13
LPoS	-1.36	-0.35	-0.36	-0.34	7.42
N/A	-14.15	6.34	6.90	7.93	0.80
PoA	0.74	-0.35	-0.36	-0.34	-0.13
PoB/PoS	0.74	-0.35	-0.36	-0.34	-0.13
POBh	-1.36	2.83	-0.36	-0.34	-0.13
PoC	-2.36	-0.61	2.97	1.29	-0.23
Pol	-1.36	-0.35	-0.36	-0.34	7.42
PoP	0.74	-0.35	-0.36	-0.34	-0.13
PoP/PoV/PoQ	0.74	-0.35	-0.36	-0.34	-0.13
PoPP	-1.36	-0.35	-0.36	2.91	-0.13
PoS	3.74	-1.77	-0.76	-2.82	-0.88
PoS/LPoS	-1.36	-0.35	-0.36	2.91	-0.13
PoS/PoB	0.74	-0.35	-0.36	-0.34	-0.13
PoS/PoW	0.74	-0.35	-0.36	-0.34	-0.13
PoS/PoW/PoT	0.74	-0.35	-0.36	-0.34	-0.13
PoSign	0.74	-0.35	-0.36	-0.34	-0.13
PoST	0.74	-0.35	-0.36	-0.34	-0.13
PoW	4.54	-2.39	-1.69	-3.12	0.63
PoW and PoS	-1.36	2.83	-0.36	-0.34	-0.13
PoW/HiPoS	0.74	-0.35	-0.36	-0.34	-0.13
PoW/nPoS	-1.36	-0.35	-0.36	2.91	-0.13
PoW/PoM/PoSII	0.74	-0.35	-0.36	-0.34	-0.13
PoW/PoS	9.09	-3.38	-4.69	-4.72	-2.43
PoW/PoS	-0.44	-0.50	-0.51	1.82	-0.19
PoW/PoS	-1.36	2.83	-0.36	-0.34	-0.13
PoWT	-1.36	2.83	-0.36	-0.34	-0.13
Proof of Authority	-1.36	-0.35	-0.36	2.91	-0.13

Table 11 Consensus algorithm - Market cap Standardized Person's residuals

	P70	P80	P90	P99	P100
536	-0.44	1.75	-0.51	-0.49	-0.19
Argon2	0.74	-0.35	-0.36	-0.34	-0.13
Argon2d	-1.36	-0.35	2.75	-0.34	-0.13
Blake	1.28	-0.61	-0.63	-0.60	-0.23
BLAKE256	-0.44	-0.50	-0.51	1.82	-0.19
Blake2b	-1.36	-0.35	-0.36	2.91	-0.13
Blake2S	1.28	-0.61	-0.63	-0.60	-0.23
C11	1.04	-0.50	-0.51	-0.49	-0.19
Counterparty	-1.92	4.00	-0.51	-0.49	-0.19
CryptoNight	0.91	-0.39	-0.45	-0.34	-0.49
CryptoNight-Lite	0.74	-0.35	-0.36	-0.34	-0.13
CryptoNight-V7	-1.92	-0.50	1.69	-0.49	5.16
Curve25519	0.74	-0.35	-0.36	-0.34	-0.13
Dagger	1.28	-0.61	-0.63	-0.60	-0.23
Dagger-Hashimoto	-1.15	1.23	1.17	-0.60	-0.23
DPoS	-0.72	-0.21	-0.27	0.83	1.84
Equihash	-2.62	0.75	-0.27	2.80	1.84
Ethash	-0.99	-0.11	-1.15	0.98	4.37
Groestl	-1.67	-0.71	2.39	0.94	-0.27
HybridScriptHash256	0.74	-0.35	-0.36	-0.34	-0.13
Keccak	-0.44	-0.50	-0.51	1.82	-0.19
Leased POS	-1.36	-0.35	-0.36	-0.34	7.42
Lyra2RE	0.42	0.89	-0.73	-0.69	-0.27
Lyra2REv2	-0.44	-0.50	-0.51	1.82	-0.19
Lyra2Z	-1.15	-0.61	1.17	1.29	-0.23
M7 POW	-1.36	-0.35	-0.36	2.91	-0.13
M7M	0.74	-0.35	-0.36	-0.34	-0.13
Mars	-1.36	-0.35	2.75	-0.34	-0.13
Momentum	0.74	-0.35	-0.36	-0.34	-0.13
Multiple	-0.40	1.10	-0.61	0.35	-0.53
N/A	-14.72	5.51	7.68	8.63	1.34
NeoScript	1.34	-1.00	0.07	-0.97	-0.38
NIST5	0.60	-1.00	1.18	-0.97	-0.38
Ouroboros	-1.36	-0.35	-0.36	-0.34	7.42
Pascal	-1.36	-0.35	2.75	-0.34	-0.13
PHI1612	-0.44	1.75	-0.51	-0.49	-0.19
PoS	2.45	-0.06	-1.73	-1.58	-0.83
POS 2.0	0.74	-0.35	-0.36	-0.34	-0.13
POS 3.0	-0.62	-0.71	0.83	0.94	-0.27
Progressive-n	0.74	-0.35	-0.36	-0.34	-0.13
Proof-of-BibleHash	-1.36	2.83	-0.36	-0.34	-0.13
Quark	1.64	-1.33	-0.53	-0.42	-0.51
QuBit	1.81	-0.87	-0.89	-0.84	-0.33
Scrypt	7.58	-2.54	-3.05	-5.06	-2.16
Scrypt-n	1.81	-0.87	-0.89	-0.84	-0.33
ScryptOG	-1.36	2.83	-0.36	-0.34	-0.13
SHA-256	-1.63	1.25	0.07	0.18	2.30
SHA-512	0.74	-0.35	-0.36	-0.34	-0.13
SHA256	3.58	-1.64	-2.14	-1.52	-0.30
SHA256D	2.21	-0.48	-1.37	-1.29	-0.51
SHA3	-0.44	-0.50	-0.51	1.82	-0.19
Shabal256	-1.36	-0.35	2.75	-0.34	-0.13
Skein	1.28	-0.61	-0.63	-0.60	-0.23
SkunkHash	0.74	-0.35	-0.36	-0.34	-0.13
SkunkHash v2 Raptor	0.74	-0.35	-0.36	-0.34	-0.13
Stanford Folding	-1.36	2.83	-0.36	-0.34	-0.13
Time Travel	-0.44	-0.50	1.69	-0.49	-0.19
VeChainThor Authority	-1.36	-0.35	-0.36	2.91	-0.13
Whirlpool	0.74	-0.35	-0.36	-0.34	-0.13
X11	6.65	-2.34	-3.68	-3.72	-0.87
X11Evo	0.74	-0.35	-0.36	-0.34	-0.13
X11GOST	-1.36	2.83	-0.36	-0.34	-0.13
X13	2.71	-1.76	-0.34	-1.68	-0.87
X14	0.74	-0.35	-0.36	-0.34	-0.13
X15	1.51	0.00	-1.09	-1.03	-0.41
XEVAN	0.74	-0.35	-0.36	-0.34	-0.13
XG Hash	0.74	-0.35	-0.36	-0.34	-0.13

Table 12 Encrypted algorithm - Market cap Standardized Person's residuals

	Cl.1	Cl.2	Cl.3	Cl.4	Cl.5
K-means	0.63	0.34	0.18		
Hist DAWass	0.27	0.86	0.16	0.45	0.62
TADPole	0.18	0.26	0.27		

Table 13 Heavy-tail cryptocurrencies, percentage of allocation on the clusters

SYM	TradDays	nonTradDays	AlphaP	Sd.P	AlphaN	Sd.N	SYM	TradDays	nonTradDays	AlphaP	Sd.P	AlphaN	Sd.N
1 007		365					BIC	200	166	1.9921	0.0724	6.3787	0.4032
2 1337	267	99	4.6784	0.2308	3.1925	0.2072	BIGUP	212	156	1.5403	0.2497	4.3769	0.3540
3 1CR	366	366	3.1808	0.1542	3.2554	0.1731	BILL	106	166	3.5035	0.1170	3.5258	0.4467
4 1ST	365	1	2.3749	0.1011	3.6957	0.2004	BIO	333	33	3.4302	0.1646	3.2460	0.1846
5 1WO	74	292	1.8122	0.0611	1.8961	0.0652	BIOB	86	280	5.9623	0.2715	6.2660	0.9309
6 2015	69	297	2.1282	0.0612	5.4682	0.8763	BIOES	178	188	2.3139	0.0929	2.0955	0.0850
7 2BACCO	148	218	1.6282	0.0439	1.6574	0.0518	BIPCC	265	101	4.4162	0.2410	3.2113	0.1721
8 2GVN	366	366	2.4278	0.1022	3.0640	0.1578	BIS	346	20	2.2119	0.0914	3.2243	0.0888
9 300	275	91	1.7354	0.0534	2.2747	0.0925	BIT	169	157	3.5119	0.1533	3.0459	0.1734
10 32BIT	1	366					BIT16	139	227	2.5644	0.0888	2.7775	0.1380
11 365	132	234					BIT8	366	3	3.3003	0.1660	2.3790	0.1045
12 404	148	218					BITOK	318	4	4.8884	0.2533	9.2627	0.1139
13 42	364	2	4.3206	0.2489	3.2916	0.1671	BITFS	90	276	2.4635	0.1799	6.0417	0.2829
14 611	25	341					BITUSD	366	2	2.3779	0.0880	2.5765	0.1433
15 608	227	139	2.1440	0.0811	2.0570	0.0818	BITZ	5	366	3.1792	0.1541	3.2890	0.1774
16 888	134	232	1.6563	0.1476	2.2782	0.1972	BKC	366	366	3.5313	0.1562	3.2644	0.1758
17 8BIT	334	32	1.7509	0.0541	1.8518	0.0648	BLAS	227	277	2.7773	0.1172	2.4628	0.1254
18 ABC	9	357					BLAZR	302	64	1.9965	0.0739	3.1079	0.1554
19 ABY	366	366	3.7879	0.2023	3.9318	0.2210	BLC	33	333	1.8852	0.0468	1.5239	0.1852
20 AC	313	105	3.2413	0.1050	4.4007	0.2507	BLOTZ	139	227	2.7773	0.1172	2.4628	0.1254
21 ACES	331	35	2.0896	0.0817	3.3651	0.1725	BLK	366	3	3.2968	0.1649	4.4746	0.2649
22 ACC*	291	75	2.5596	0.1117	2.5694	0.1200	BLOCK	366	44	4.9149	0.2789	2.3494	0.1038
23 ACID	145	221	3.1295	0.1206	4.0567	0.4160	BLRY	141	135	2.5581	0.0915	1.0820	0.0892
24 ACH	366	366	3.2371	0.1582	3.3896	0.1855	BLU	244	225	2.9784	0.1234	2.8934	0.1814
25 ACN	109	257	2.7239	0.0956	3.6475	0.4135	BLX	113	250	2.0769	0.0769	2.0525	0.0807
26 ACN	366	366					BNH	366	3	3.2988	0.1713	2.0589	0.0776
27 ACQIN	69	297	3.0579	0.1108	2.8390	0.4013	BMC	366	270	2.3640	0.0770	3.2575	0.2547
28 ACP	247	252	2.5229	0.0849	2.2654	0.1908	BNEP	96	94	1.9359	0.0690	4.5800	0.2654
29 ADA	326	40	4.5476	0.1824	7.8722	0.5255	BNS	272	94	5.2378	0.3074	3.4386	0.1838
30 ADC	132	234	7.6750	0.3749	5.4270	0.6324	BNT	366	276	2.5247	0.1318	2.2517	0.1769
31 ADEN	305	61	1.5760	0.0396	1.5118	0.0412	BOL	231	94	2.7200	0.1251	3.0840	0.1566
32 ADL	113	253	3.3537	0.1316	2.3849	0.2042	BOAT	172	244	2.3313	0.0990	2.3641	0.1056
33 ADN	366	366	17.1788	1.1412	3.2603	0.1760	BOB	366	356	3.5167	0.1520	3.5738	0.1949
34 ADST	365	1	2.9668	0.1463	2.4323	0.1047	BOLI	104	248	4.0848	0.1735	3.9420	0.4161
35 ADT	366	366	2.8627	0.1429	2.2766	0.0912	BOMB	152	214	3.2219	0.1244	3.4182	0.3527
36 ADX	366	366	3.2289	0.1675	3.5532	0.1864	BOB	138	33	3.1675	0.1551	3.3814	0.1846
37 ADZ	356	10	3.3311	0.1708	3.9878	0.3913	BOSON	231	225	3.9233	0.4359	1.7511	0.2265
38 AE	366	366	4.2636	0.2393	6.4882	0.4098	BOSS	33	33	3.2742	0.2044	2.0669	0.0782
39 14EC	366	227	1.7443	0.0617	1.4074	0.0609	BOSS	33	33	3.2742	0.2044	2.0669	0.0782
40 AEON	366	366	4.6230	0.2588	3.6646	0.2044	BPL	314	53	3.1668	0.0359	2.4497	0.3164
41 AERO	366	366	3.2124	0.1572	3.2909	0.1767	BP	93	366	3.2742	0.2044	2.0669	0.0782
42 AGES	70	296	3.9168	0.1739	3.4369	0.2514	BR	366	366	3.2742	0.2044	2.0669	0.0782
43 AGS	366	366	17.5597	1.1709	3.3789	0.1846	BRAIN	366	366	3.2742	0.2044	2.0669	0.0782
44 AHT	206	140	1.5248	0.0368	2.1052	0.0866	BR	366	366	3.2742	0.2044	2.0669	0.0782
45 AIB	366	30	2.5716	0.1198	2.4439	0.1396	BRD	156	210	3.1763	0.1520	3.5738	0.1949
46 AIR	328	38	2.3742	0.1008	4.3917	0.2528	BRIT	366	366	3.2742	0.2044	2.0669	0.0782
47 ALC	158	208	1.5814	0.0413	1.5694	0.0439	BRK	366	366	3.2742	0.2044	2.0669	0.0782
48 ALEX	362	4					BRN	366	366	3.2742	0.2044	2.0669	0.0782
49 ALF	366	366	17.5410	1.1785	3.2448	0.1727	BRONZ	3	363	3.4389	0.1818	10.6880	0.1704
50 ALIS	71	295	5.0383	0.2206	5.4118	0.7924	BRB	366	366	3.4389	0.1818	10.6880	0.1704
51 ALN	366	366	17.3853	1.1557	3.3821	0.1854	BRB	366	366	3.4389	0.1818	10.6880	0.1704
52 ALT	72	294	2.4712	0.0812	2.2377	0.2008	BSC	366	276	3.5768	0.4416	1.2156	0.3747
53 ALTCOM	66	230	2.0022	0.1238	3.3861	0.1948	BSC	366	276	2.6655	0.0808	3.6762	0.1957
54 ALTOCAR	14	352	2.7999	0.1184	2.3321	0.1146	BST	157	2	3.6449	0.2538	7.4315	0.1818
55 AM	366	366	3.8997	0.2104	3.7684	0.2087	BSTAR	319	319	2.1344	0.0606	3.9558	0.7632
56 AMB	366	366	3.8488	0.1552	4.9419	0.7320	BSTY	7	359	3.8488	0.1552	4.9419	0.7320
57 AMBER	92	274	17.8368	1.1905	3.3762	0.1844	BT1	1	366	3.8488	0.1552	4.9419	0.7320
58 AMC	366	160	1.3017	0.0236	1.3503	0.0261	BT2	366	366	3.8488	0.1552	4.9419	0.7320
59 AMIS	366	366	3.3020	0.1771	3.4523	0.1747	BT4	366	2	2.5761	0.1171	3.1919	0.1612
60 AMM	357	9	4.2680	0.2319	3.5742	0.2004	BTB	366	1	2.1977	0.0893	4.5224	0.2583
61 AMP	366	366	3.8179	0.4130	6.9576	0.8262	BTB	366	366	3.8179	0.4130	6.9576	0.8262
62 AMS	134	232	17.3513	1.1562	3.2646	0.1758	BTCA	254	112	1.6797	0.0437	2.0036	0.0901
63 AMX	366	366	1.8773	0.0605	3.3315	0.1026	BTB	267	97	2.1277	0.1213	1.1861	0.0926
64 AMY	193	84	2.0161	0.0730	2.0999	0.0984	BTCE	242	282	1.7951	0.0540	1.7447	0.0610
65 ANAL	100	366	3.2389	0.1583	3.3762	0.1844	BTCH	330	36	3.0593	0.1486	3.3788	0.1803
66 ANC	366	366	3.2389	0.1583	3.3762	0.1844	BTCH	330	36	3.0593	0.1486	3.3788	0.1803
67 AND	366	366	3.2389	0.1583	3.3762	0.1844	BTCH	330	36	3.0593	0.1486	3.3788	0.1803
68 ANGL	25	341	2.2046	0.0831	1.9628	0.0771	BTCH	330	65	3.4561	0.1585	3.6976	0.2403
69 ANI	366	366	2.7336	0.1354	2.7108	0.1204	BTCH	330	157	4.4999	0.2078	2.6905	0.1845
70 ANT	366	366	2.7336	0.1354	2.7108	0.1204	BTCH	330	157	4.4999	0.2078	2.6905	0.1845
71 ANTI	87	279	4.1914	0.1721	4.3382	0.7177	BTCH	330	249	3.8227	0.2064	3.4416	0.1825
72 APC	313	35	3.3103	0.1541	2.9661	0.1486	BTB	156	208	2.7158	0.0995	3.1668	0.3064
73 APX	366	366	17.2295	1.1475	3.2779	0.1758	BTB	156	187	3.1955	0.1227	3.1951	0.1723
74 APT	303	65	4.2786	0.2360	2.3921	0.1058	BTE	366	366	3.1955	0.1227	3.1951	0.1723
75 AQUA	342	24	6.6617	0.4325	2.7867	0.1347	BTG	366	366	3.1955	0.1227	3.1951	0.1723
76 ARB	109	257	2.8063	0.1343	1.7543	0.0555	BTLC	183	8	3.7478	0.1355	2.5738	0.2949
77 ARS	366	366	3.2389	0.1583	3.3762	0.1844	BTCH	330	36	3.0593	0.1486	3.3788	0.1803
78 ARC*	134	232	2.5500	0.0896	2.7047	0.2083	BTM*	366	366	3.2389	0.1583	3.3762	0.1844
79 ARCH	366	366	17.7345	1.1833	3.3795	0.1847	BTM	366	366	17.7345	1.1833	3.3795	0.1847
80 ARCO	366	366	17.7345	1.1833	3.3795	0.1847	BTM	366	366	17.7345	1.1833	3.3795	0.1847
81 ARD	366	366	17.7345	1.1833	3.3795	0.1847	BTM	366	366	17.7345	1.1833	3.3795	0.1847
82 ARD*	366	366	17.7345	1.1833	3.3795	0.1847	BTM	366	366	17.7345	1.1833	3.3795	0.1847
83 ARENA	3	363	3.3587	0.1660	2.8737	0.1463	BTSE	274	10	3.2942	0.1576	2.4005	0.1135
84 ARE	366	366	3.3587	0.1660	2.8737	0.1463	BTSE	274	10	3.2942	0.1576	2.4005	0.1135
85 ARGUS	71	295	2.6901	0.0926	3.5429	0.4427	BTX	366	366	3.3587	0.1660	2.8737	0.1463
86 ARI	363	3	3.9805	0.1452	2.3729	0.1023	BTX	366	366	3.9805	0.1452	2.3729	0.1023
87 ARK	366	366	3.4212	0.3213	2.1323	0.1023	BTX	366	366	3.4212	0.3213	2.1323	0.1023
88 ARM	169	297	2.6120	0.0913	2.9191	0.2612	BTZ	172	194	2.0474	0.0752	1.9348	0.0713
89 ARN	366	366	2.6120	0.0913	2.9191	0.2612	BTZ	172					

SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	
151 CRAB	4	362					EGG							GHOU	366						
152 CRE	323	43	3.5527	0.1823	2.4104	0.1082	EGO	90	276	1.9916	0.0721	2.0071	0.0757	GIFT	155	211	3.6036	0.1841	2.2061	0.0936	
153 CRE	3	363					EGN	366						GIG	112	254	2.1438	0.0807	3.7168	0.1185	
154 CREA	365	1	3.7853	0.2048	3.7345	0.2033	EKO	358	8	2.4857	0.1260	3.1982	0.1459	GIM	101	265	1.9290	0.0652	1.8589	0.0673	
155 CRED	218	148	1.6703	0.0457	1.8224	0.0669	ELC	3	363					GIN	189	177	2.5388	0.1240	4.9526	0.2715	
156 CREDO	217	149	4.5326	0.2670	1.7929	0.0574	ELE	257	109	3.3279	0.1303	2.7450	0.2545	GIO	366	3.2371	0.1582	3.3896	0.1655		
157 CREVA	176	190	1.8830	0.0626	3.3937	0.1852	ELITE	275	91	3.8637	0.1665	1.7584	0.0906	GIZ	8	358					
158 CRM	8	358					ELIX	365	1	4.1404	0.2430	3.9500	0.2091	GLC	121	245	2.8220	0.1022	12.7945	1.7024	
159 CRPS	137	229	1.8566	0.0606	1.9325	0.0724	ELIJA	366	1	3.2144	0.1646	6.9113	0.6331	GLD	366	3.4457	0.1876	2.5792	0.1298		
160 CRTM	66	300	1.8692	0.0590	1.7138	0.0585	ELM	5	361					GLOBE	366	17.3513	1.5662	3.2646	0.1758		
161 CRW	366	3	3.9985	0.1921	3.9410	0.2174	ELP	265	101	2.7552	0.1319	1.8334	0.0606	GLT	366	17.3027	1.1528	3.4022	0.1864		
162 CRX	127	239	3.7271	0.1544	5.1096	0.5592	ELS	366	3	3.2007	0.1560	3.3692	0.1839	GLX	366	3.2047	0.1599	3.3855	0.1852		
163 CRYPT	41	325	4.3158	0.2241	2.6517	0.1382	ELTCCN	31	335	2.2604	0.0914	2.3982	0.1054	GLYPH	366	19.5093	1.3023	3.2641	0.1768		
164 CS	120	246	1.5074	0.1406	3.2004	0.3276	ELTCON	136	230	1.7207	0.0903	2.1801	0.0930	GMC	366						
165 CSC	106	260	2.5077	0.0944	2.7532	0.1664	EMB	233	133	5.4007	0.2627	4.9905	0.4138	GML	60	306	1.7708	0.0413	3.9044	0.7044	
166 CSH	366	366					EMC	366	33.945	0.1810	3.5596	0.1852	GMX	37	329	2.8848	0.1005	1.9455	0.2527		
167 CSMC	1	365					EMC2	366	3.5072	0.1833	6.8833	0.4997	CNJ	366							
168 CSNO	354	12	3.4536	0.1804	6.5587	0.4132	EMD	312	54	3.1947	0.1564	3.8996	0.2230	GNO	366	2.7229	4.2857	0.2449			
169 CST	164	202	1.8470	0.0618	2.1118	0.0833	EMPC	5	361					GNT	366	3.1522	0.1553	5.8534	0.3679		
170 CTC	366	366					ENV	125	241	1.7795	0.0579	1.8758	0.0644	COAT	38	328	3.4197	0.1281	7.8831	0.2944	
171 CTCIC	2	364					ENE	150	216	1.9379	0.0658	1.8999	0.0705	GOLOS	366	2.4932	0.1095	2.3772	0.1026		
172 CTR	153	350					ENG	366	3.5254	0.1832	2.3306	0.1003	GOOD	8	358	2.3361	0.0950	2.5947	0.1230		
173 CTR	44	322	4.9841	0.2308	2.9608	0.2378	ENJ	366	2.2225	0.0909	2.4529	0.1054	GOON	366	216	1.7489	0.0519	1.7674	0.0610		
174 CTT	44	322	2.3549	0.0924	2.1724	0.0954	ENRG	30	66	3.4594	0.1757	3.4187	0.1855	GOT	117	249	5.0390	0.2244	4.2824	0.5045	
175 CTX	351	15	2.2049	0.0881	2.3336	0.0997	ENT	153	13	1.7663	0.0569	2.4514	0.1830	GOTX	152	214	1.6679	0.0393	5.8950	0.6573	
176 CUBE	5	361					ENTER	97	269	3.1334	0.1288	5.2898	0.6959	GP	5	361					
177 CURE	366	366	5.6559	0.3459	2.3956	0.1029	ENTRC	35	351	2.6781	0.1057	1.9029	0.0846	GPL	273	93	4.6758	0.2593	3.0698	0.1611	
178 CVC	366	3.7356	0.1714	5.5330	0.3452	EOC	178	18	2.1452	0.0820	5.3679	0.3979	GRL	159	207	4.8221	0.2174	2.2073	0.1599		
179 CWT	211	155	2.0915	0.0832	5.1499	0.2979	EOS	366	2.9476	0.1424	3.5227	0.1886	GRAM	302	64	2.6668	0.1092	3.1017	0.1822		
180 CWXT	214	152	1.6604	0.0461	3.3561	0.1857	EPY	1	365					GRAV	5	361					
181 CXC	366	3.1982	0.1547	3.3433	0.1830	EQ	366	366	3.1477	0.1519	3.2729	0.1764	GRD	366	41.780	0.2264	6.2324	0.2018			
182 CXT	73	293	1.9194	0.0689	2.0109	0.0737	EQM	2	364					GRE	148	218	4.1504	0.1816	2.6072	0.1993	
183 CVC	65	301	4.7144	0.2000	6.2967	1.1558	EQT	2	364					GREXIT	173	193	3.5836	0.1869	2.1772	0.0990	
184 CYG	268	1.7375	0.0400	2.4737	0.3282	EQUAL	8	9	366					GRG	366						
185 CYP	4	362					ERC	286	80	6.1071	0.3657	2.3138	0.1005	GRM	366	3.2284	0.1576	3.3018	0.1787		
186 CZZ	366	3.1896	0.1545	3.3629	0.1849	EREAL	140	226	2.7733	0.1270	1.5719	0.0437	GROW	366	2.9696	0.1472	4.0881	0.2258			
187 DAI	274	92	1.8714	0.0519	1.7027	0.0767	ERO	13	366	3.8012	0.2166	3.1322	0.1951	HNH	366	366					
188 DANK	366	3.2217	0.1567	3.3715	0.1846	ERR	97	269	2.0028	0.0713	2.0239	0.0790	GRW	302	64	3.3438	0.1653	5.3645	0.3398		
189 DARK	244	122	3.5896	0.1622	7.0147	0.4697	ESP	366	241	125	3.3885	0.1702	4.6457	0.2804	CSM	202	140	1.7898	0.0540	4.2266	0.2617
190 DAS	366	3.3554	0.1722	2.3559	0.1013	EST	366	366	2.6057	0.1162	3.0353	0.1539	GSV	134	2	2.1669	0.0787	1.9936	0.0822		
191 DATA	366	3.9599	0.2225	7.7790	0.4931	ETBS	337	23	3.2835	0.0929	1.9449	0.0787	GUE	366	17.1052	1.1388	3.2823	0.1771			
192 DASH	366	4.3006	0.2440	3.1756	0.1608	ETBT	30	366	4.6969	0.2689	6.4972	0.4132	GUN	296	70	2.2617	0.0935	2.2820	0.0945		
194 DAXX	316	50	2.1545	0.0829	3.7090	0.2066	ETC	366	250	116	4.0573	0.1934	2.4837	0.1123	9.4410	0.0704	3.4102	0.1123			
195 DAY	61	2.9604	0.1473	3.7563	0.1889	ETD	366	366	9.3143	0.5969	3.1858	0.1667	GUP	366	3.2365	0.1610	3.6725	0.2032			
196 DB	24	362					ETH	366	366	2.4002	0.1036	2.6374	0.1227	GVT	366	3.3465	0.1744	6.3068	0.5732		
197 DBET	364	2	2.9855	0.1492	4.7186	0.2705	ETHB	12	354	2.4002	0.1036	2.6374	0.1227	GVT	366	3.3465	0.1744	6.3068	0.5732		
198 DBIC	142	2.1608	0.0833	2.4508	0.1083	ETHD	366	366	2.2225	0.0909	2.4529	0.1054	GOON	366	216	1.7489	0.0519	1.7674	0.0610		
199 DBIC	71	295	4.9037	0.2142	2.6426	0.2817	ETHOS	349	17	2.3668	0.1005	3.4310	0.1807	GXC*	366	3.2616	0.1568	3.1018	0.1794		
200 DBIX	366	2.2985	0.0990	3.6646	0.1913	ETHS	117	249	3.1656	0.1209	3.4404	0.3638	GXS	366	3.6604	0.1967	3.0021	0.1480			
201 DBTC	366	207	3.0192	0.1158	3.5951	0.3249	ETI	366	366	2.1857	0.1392	3.3901	0.1743	HAC	366	1.88	1.6299	0.0454	1.5397		
202 DCC	63	303	2.7977	0.4237	2.3855	0.0743	ETN	366	1	2.8577	0.1392	3.3901	0.1743	HAC	366	4	3.62				
203 DCC	78	288	2.5430	0.1385	3.1079	0.3914	ETP	366	366	2.1857	0.1392	3.3901	0.1743	HAC	366	19	2.6164	0.1223	4.710	0.1812	
204 DCN	340	26	3.7507	0.2068	4.0694	0.2233	ETI	366	30	3.8194	0.2119	3.2755	0.1655	HALLO	94	272	2.7164	0.0946	2.6915	0.2781	
205 DCR	366	28	2.4317	0.1033	4.1005	0.2350	EUC	75	291	7.0436	0.3263	1.7968	0.1662	HAMS	64	302	6.1697	0.2088	4.5863	0.6021	
206 DCRE	5	361					EVC	50	316	0.9955	0.0721	6.8524	0.2025	HAZ	366	366					
207 DCY	353	13	3.6505	0.1943	6.0855	0.3791	EVN	79	366	3.2132	0.1565	3.2921	0.1779	HSN	230	136	3.3003	0.1623	3.2411	0.1745	
208 DDF	138	228	1.9936	0.0692	3.7948	0.2210	EVIL	716	287	4.9994	0.2159	6.1123	0.0660	HTI	282	84	2.9642	0.1456	2.8571	0.1369	
209 DEA	224	142	2.7254	0.1304	3.2870	0.1008	EVN	237	129	3.1402	0.1600	2.5287	0.1118	HCC	198	168	3.0119	0.1205	2.5570	0.1669	
210 DEEP	19	347					EVX	366	116	2.9302	0.1447	2.2542	0.0915	HDC	91	275	3.4126	0.1336	2.3523	0.2138	
211 DEM	172	194	11.8288	0.6242	6.4158	0.6718	EXR	250	366	2.9302	0.1447	2.2542	0.0915	HDC	91	275	3.4126	0.1336	2.3523	0.2138	
212 DES	6	360					EXC	366	366	1.4636	0.1642	3.3805	0.1848	HILL	366	3.6060	0.1901	3.4472	0.1834		
213 DETH	98	268	3.4755	0.1365	3.5514	0.4194	EXCL	366	366	2.9302	0.1447	2.2542	0.0915	HDC	91	275	3.4126	0.1336	2.3523	0.2138	
214 DEUR	366	3.4755	0.1365	3.5514	0.4194	EXCL	366	366	2.9302	0.1447	2.2542	0.0915	HDC	91	275	3.4126	0.1336	2.3523	0.2138		
215 DFS	6	360	3.7360	0.1628	2.7560	0.1420	EXT	162	204	5.0786	0.2335	6.0371	0.6449	HIRE	72	294	4.0490	0.6654	6.1812	0.2789	
216 DFT	365	1	3.5658	0.1771	5.0238	0.3222	EXN	282	84	1.4958	0.0353	19.0197	1.3861	HKG	366	16.8557	1.1212	3.3739	0.1843		

SYM	TradDays	nonTradDays	AlphaP	Sd.P	AlphaN	Sd.N	SYM	TradDays	nonTradDays	AlphaP	Sd.P	AlphaN	Sd.N	SYM	TradDays	nonTradDays	AlphaP	Sd.P	AlphaN	Sd.N	
301 ICON	209	157	13.381	0.7402	3.7424	0.2567	MAD*	168	198	3.6625	0.1527	2.9963	0.2535	NTM	344	22	3.8359	0.1962	2.2869	0.1027	
302 KOS	186	162	2.241	0.0788	2.1467	0.1251	MAID	366	198	4.0082	0.2101	3.5522	0.2011	NTD	291	75	1.9623	0.0387	2.0353	0.183	
303 ICX	366	366	3.6887	0.1951	3.8527	0.2150	MAN	225	141	2.8115	0.1332	2.6089	0.1196	NTRN	366	196	2.3372	0.0981	2.4366	0.1071	
304 IEC	4	362	3.1573	0.1404	3.4546	0.2153	MANA	366	1	3.1776	0.1572	3.4786	0.1879	NTRN	196	170	3.7824	0.2057	1.9953	0.0736	
305 IETH	287	79	3.1573	0.1404	3.4546	0.2153	MANA	366	1	3.1776	0.1572	3.4786	0.1879	NTRN	196	170	3.7824	0.2057	1.9953	0.0736	
306 IFC	366	366	3.2176	0.1568	3.3108	0.1794	MAR	331	35	3.0370	0.1470	2.3250	0.1004	NURK	54	312	2.2728	0.0685	2.3581	0.2964	
307 IFLT	947	179	3.3077	0.1648	3.4084	0.1824	MARV	337	37	3.6163	0.1918	3.6340	0.1963	NURB	366	5	5.9285	0.3594	4.0531	0.2438	
308 IFT	341	275	6.0357	0.2776	4.9643	0.6517	MARV	118	248	2.7085	0.0954	4.0401	0.4532	NUM	112	254	3.5222	0.1410	4.7383	0.5512	
309 ILC	143	223	2.6403	0.0941	2.6211	0.2059	MARX	107	259	3.2467	0.1252	3.6180	0.3947	NVC	366	3	3.3928	0.1759	3.3302	0.1732	
310 ILT	366	366	1.7438	1.1618	3.3602	0.1832	MARVY	366	366	1.7438	1.1618	3.3602	0.1832	MARVY	366	366	1.7438	1.1618	3.3602	0.1832	
311 IMPL	366	366	2.4757	0.0993	2.3935	0.1157	MAT	43	323	2.1156	0.0759	1.9952	0.0813	NXE	6	360	2.5425	0.1143	3.1485	0.1584	
312 IMPCH	151	366	17.7255	1.1827	3.3818	0.1849	MAT*	175	191	2.6979	0.0964	3.1162	0.2828	NXS	366	2	4.6745	0.2716	3.6593	0.1966	
313 IMPS	157	215	5.0956	0.2840	2.8091	0.1439	MAX	325	41	2.9423	0.1010	3.3805	0.1202	NXT	366	2	4.6745	0.2716	3.6593	0.1966	
314 IMS	101	366	1.6338	0.0348	5.8123	0.8253	MAX	114	252	5.8338	0.2594	10.6116	1.4020	NXTI	366	366	3.1815	0.1546	3.3897	0.1849	
315 IMX	366	366	16.8557	1.2112	3.3739	0.1843	MBC	366	366	3.2085	0.1566	3.2635	0.1782	NXTTY	366	366	3.2176	0.1568	3.3108	0.1794	
316 IN	71	295	5.4535	0.2419	2.3259	0.4297	MBI	198	168	1.8013	0.0580	1.6594	0.0468	NVAN	337	29	3.1074	0.1554	2.3717	0.1017	
317 INC	213	215	2.7907	0.1350	4.6530	0.2650	MBIT	53	313	1.8640	0.0465	4.8395	0.8378	NVC	10	356	2.4736	0.1078	5.9811	0.3723	
318 INCHT	366	366	3.0197	0.1538	3.0250	0.1260	MBRS	328	38	2.1154	0.1539	2.4816	0.1114	OAX	366	1	2.3804	0.0964	2.2634	0.0996	
319 IND	157	209	2.6943	0.0995	2.7850	0.2048	MCAP	296	80	2.6335	0.1075	2.3787	0.1187	OBITS	50	316	3.3928	0.1759	3.3302	0.1732	
320 INDI	338	28	2.7290	0.1307	4.9210	0.2837	MCAR	348	18	2.5105	0.1183	4.5196	0.2470	OBS	6	360	2.5425	0.1143	3.1485	0.1584	
321 INF3	103	265	1.6954	0.0388	3.1257	0.3025	MCI	279	87	2.1321	0.0791	2.9156	0.1510	OCEAN	9	357	4.9290	0.2822	4.0155	0.2525	
322 INN	365	1	3.1170	0.1581	3.5932	0.1907	MCO	366	366	4.3202	0.2441	4.8155	0.2836	OCL	276	90	3.3643	0.1672	3.6259	0.2038	
323 INSANE	133	355	3.11	2.0291	0.0550	2.8234	0.4559	MCRN	211	155	4.1010	0.1880	3.4875	0.2576	OCL	336	30	2.7368	0.1433	3.4202	0.1635
324 INSN	366	366	3.4213	0.1771	4.5053	0.1831	MDT	366	366	3.9531	0.1444	4.0243	0.2236	OCTO	366	1	1.9623	0.0387	2.0353	0.183	
325 INV	126	240	1.7869	0.0542	1.8191	0.0658	MDC	366	366	16.9092	1.1250	3.4089	0.1870	ODN	365	1	3.5030	0.1840	4.5386	0.2630	
326 IOC	366	366	3.4213	0.1771	4.5053	0.1831	MDT	366	366	16.9092	1.1250	3.4089	0.1870	ODN	365	1	3.5030	0.1840	4.5386	0.2630	
327 ION	366	366	2.6489	0.1209	2.5979	0.1191	MEC	201	165	1.6796	0.1807	3.8532	0.2370	OC	366	275	8.7380	0.4222	3.9787	0.5438	
328 IOP	366	366	3.7541	0.1977	5.3588	0.3324	MED	189	177	2.9489	0.1351	3.0321	0.1617	OLDSP	366	366	17.7345	1.1833	3.3795	0.1847	
329 IOT	366	366	3.9811	0.2180	3.0741	0.1550	MEDI	19	47	2.1155	0.0795	1.9956	0.0721	OLVMP	124	242	5.7187	0.2671	16.1760	0.2552	
330 IOU	366	366	3.2115	0.1564	3.3808	0.1848	MEGA	366	366	17.2039	1.1458	3.2641	0.1757	OMA	6	360	2.5425	0.1143	3.1485	0.1584	
331 IPC	366	366	17.7296	1.1830	3.3971	0.1860	MEME	366	366	3.4285	0.1753	3.8939	0.2194	OMI	57	309	3.7511	0.1488	4.5825	0.7313	
332 IRL	354	366	3.1776	0.1572	3.4786	0.1879	MER	366	366	3.1776	0.1572	3.4786	0.1879	MER	366	366	3.1776	0.1572	3.4786	0.1879	
333 ISL	366	366	3.1776	0.1572	3.4786	0.1879	MER	366	366	3.1776	0.1572	3.4786	0.1879	MER	366	366	3.1776	0.1572	3.4786	0.1879	
334 ITT	214	152	2.1173	0.0788	2.0078	0.0785	MET	366	366	17.5738	1.1749	3.3625	0.1828	ORO	366	366	3.1815	0.1546	3.3897	0.1849	
335 ITZ	133	366	1.6338	0.0348	5.8123	0.8253	METAL	71	95	3.1776	0.1572	3.4786	0.1879	METAL	366	366	3.1776	0.1572	3.4786	0.1879	
336 IYZ	214	213	2.0155	0.2722	0.0497	0.8010	MG	366	366	3.2132	0.1565	3.9291	0.1719	OPAL	61	305	4.0380	0.1641	4.2255	0.6842	
337 IW	366	366	3.1776	0.1572	3.4786	0.1879	MGO	366	366	3.1776	0.1572	3.4786	0.1879	OPC	366	366	3.1776	0.1572	3.4786	0.1879	
338 IWT	21	345	2.1707	0.0814	1.9359	0.0742	MHC	11	355	2.3272	0.0740	1.7577	0.1142	OPC	366	366	3.1776	0.1572	3.4786	0.1879	
339 IXC	366	366	3.1776	0.1572	3.4786	0.1879	MHC	11	355	2.3272	0.0740	1.7577	0.1142	OPC	366	366	3.1776	0.1572	3.4786	0.1879	
340 IXT	366	366	3.1776	0.1572	3.4786	0.1879	MHC	11	355	2.3272	0.0740	1.7577	0.1142	OPC	366	366	3.1776	0.1572	3.4786	0.1879	
341 J	302	65	3.3829	0.1698	3.2066	0.1697	MIL	366	366	3.2122	0.1564	3.2968	0.1783	ORL	364	2	3.1091	0.1563	4.9902	0.2942	
342 JANE	111	254	4.2527	0.1821	4.2870	0.4795	MILO	88	278	19.9998	1.0470	5.8806	0.8024	ORLY	105	261	5.1370	0.7313	5.4736	0.2448	
343 JBS	366	366	3.1518	0.1462	3.1806	0.1756	MIN	366	366	3.1518	0.1462	3.1806	0.1756	MIN	366	366	3.1518	0.1462	3.1806	0.1756	
344 JET	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
345 JIF	210	156	2.3400	0.0967	2.6115	0.1222	MINT	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
346 JO	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
347 JKC	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
348 JNS	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
349 JNT	322	44	2.4959	0.1343	4.1292	0.2012	MLN	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
350 JOBS	148	218	4.6228	0.2064	4.3327	0.4376	MLS	366	366	3.2176	0.1568	3.3108	0.1794	MIN	366	366	3.2176	0.1568	3.3108	0.1794	
351 JOK	182	182	2.7573	0.1809	0.0900	0.0900	MM	135	261	1.8326	0.0575	3.9932	0.2396	OC	366	366	3.2176	0.1568	3.3108	0.1794	
352 JUDGE	366	366	17.3646	1.1572	3.3914	0.1856	MMC	366	366	3.2345	0.1584	3.3954	0.1854	PAK	226	140	3.1952	0.1291	2.5991	0.1454	
353 JWL	107	259	8.9390	0.4149	11.9890	0.1737	MMNXT	366	366	3.2345	0.1584	3.3954	0.1854	PAK	226	140	3.1952	0.1291	2.5991	0.1454	
354 KARMA	366	366	3.1518	0.1462	3.1806	0.1756	MMNXT	366	366	3.1518	0.1462	3.1806	0.1756	MIN	366	366	3.1518	0.1462	3.1806	0.1756	
355 KAT	137	230	4.6160	0.1741	4.1017	0.5519	MMNXT	366	366	3.1518	0.1462	3.1806	0.1756	MIN	366	366	3.1518	0.1462	3.1806	0.1756	
356 KAY	366	366	3.1518	0.1462	3.1806	0.1756	MMNXT	366	366	3.1518	0.1462	3.1806	0.1756	MIN	366	366	3.1518	0.1462	3.1806	0.1756	
357 KBC	169	197	2.2329	0.0960	2.6584	0.1170	MNC	321	45	2.9006	0.0629	1.8464	0.0671	PAP	202	164	2.7628	0.1001	3.6188	0.2768	
358 KBR	363	3	2.9526	0.1453	2.3968	0.1723	MND	8	358	1.9049	0.0629	1.8464	0.0671	PAP	202	164	2.7628	0.1001	3.6188	0.2768	
359 KC	137	229	2.2329	0.0960	2.6584	0.1170	MNC	321	45	2.9006	0.0629	1.8464									

SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	
451 POP							501							551							
452 POST	346	20	17.3766	1.1580	3.3866	0.1852	SH	79	287	2.7528	0.0956	2.0304	0.1881	TKS	366	39	3.4908	0.1774	2.4033	0.1083	
454 POT	366		2.1611	0.2371	3.3656	0.1753	SHADE	327	366	3.6897	0.1897	3.9712	0.2313	TMC	201	15	3.1605	0.1932	4.4858	0.2536	
455 POWER	366		3.1580	0.1557	3.3608	0.1790	SHIFT	366	366	3.2176	0.1568	3.4052	0.1867	TME	366	366	18.0885	1.2562	4.4334	0.1831	
456 PPC	366		2.5232	0.1074	3.4746	0.1926	SHLD	193	366	4.0666	0.2236	4.1247	0.2342	TNT	366	366	2.8784	0.1377	3.7680	0.2063	
457 PPP	330	36	2.7786	0.1297	3.0132	0.1509	SHRST	225	173	2.0126	0.0712	1.8861	0.0692	TODAY	231	135	1.6853	0.0475	5.3793	0.3484	
458 PPT	330	36	3.9029	0.2152	3.5398	0.1872	SHORTTY	172	141	2.8975	0.1362	1.9184	0.0700	TOK	360	6	3.8576	0.2160	3.8560	0.2067	
459 PPV	330	36	5.8895	0.3644	5.5364	0.3326	SHREK	172	141	2.8472	0.1141	4.1955	0.3690	TOKC	297	69	2.1171	0.1060	3.0003	0.1363	
460 PRE	103	73	2.3396	0.0732	1.8659	0.1555	SIB	366	366	3.4862	0.1888	2.2693	0.0918	TOM	1	365	3.0855	0.1378	3.1071	0.1800	
461 PRE	260	260	2.6783	0.0932	2.6627	0.2612	SIC	78	288	4.6795	0.1995	4.8043	0.7579	TOR	110	266	2.7616	0.1471	0.9045	2.8209	
462 PRES	346	64	3.2014	0.1606	3.2028	0.1651	SHB	301	270	2.0512	0.0748	1.9869	0.0762	TOT	366	165	1.8625	0.0433	1.7168	0.0572	
463 PRFT	367	309	1.8997	0.0641	1.8058	0.0620	SIGT	138	228	2.5115	0.0849	8.6747	0.1064	TPG	178	188	2.6255	0.0433	1.7168	0.0572	
464 PRG	51	5	2.9513	0.1442	2.1472	0.0848	SILK	172	141	3.2004	0.1556	3.3840	0.1590	TRA	1	365	1				
465 PRIME	366						SISA	173	193	1.8572	0.0611	1.8954	0.0689	TRC	33	333	1.8913	0.0476	2.1610	0.2903	
466 PRM	106	260	2.1781	0.0658	5.8157	0.7179	SJCK	101	265	1.9248	0.0628	1.7922	0.0649	TRCT	345	21	3.4291	0.1791	3.7059	0.2006	
467 PRL	302	64	3.1810	0.1549	2.5023	0.1919	SIFT	96	270	2.0512	0.0748	1.9869	0.0762	TTC	366	366	3.0509	0.1552	2.4484	0.1051	
468 PRX	154	212	2.9921	0.1368	2.1317	0.0912	SKC	364	1	2.7114	0.1276	3.6115	0.1915	TRI	332	34	2.9642	0.1425	7.9789	0.5261	
469 PRO	366	366	3.9559	0.2228	4.2803	0.2380	SKIN	361	5	4.0005	0.2165	3.0779	0.1575	TRIA	142	224	1.8078	0.0555	1.7918	0.0638	
470 PROD	7	359					SKR	349	121	4.1334	0.2291	8.6878	0.5746	TRICK	63	303	2.6097	0.0868	2.1167	0.1331	
471 PRT	366		3.66	0.2371	1.582	0.3896	0.1855	SKULL	103	263	4.4966	0.1925	5.0374	0.6729	TRIG	286	80	3.8216	0.2010	3.4757	0.1904
472 PRX	79	287	2.6924	0.0930	2.4926	0.2523	SKY	365	1	4.3058	0.2398	3.6130	0.1970	TRK	353	13	3.0949	0.1536	1.1852	0.1629	
473 PSB	302	64	3.1810	0.1549	2.5023	0.1919	SLS	245	121	2.2584	0.0909	3.8846	0.2187	TROLL	12	354	12	3.2896	0.1626	3.2028	0.1676
474 PSEUD	366		17.3513	1.1562	3.2644	0.1758	SLM	366	366	3.3510	0.1778	3.2972	0.1662	TRUMP	221	145	2.1116	0.0617	1.9300	0.0691	
475 PSI	135	231	1.8477	0.0607	1.9181	0.0702	SLR	366	366	3.4925	0.1894	3.5756	0.1884	TRUST	300	66	2.4083	0.1019	2.3363	0.1010	
476 PST	366		4.0241	0.2286	3.6730	0.1934	SLS	366	29	317	1.8572	0.0611	1.8954	0.0689	TRC	33	333	1.8913	0.0476	2.1610	0.2903
477 PSY	6	360					SLT	16	350	1.9346	0.0766	2.3547	0.0920	TRV	5	361	3.0370	0.1603	2.8205	0.1450	
478 PTA	27	339	2.7136	0.0909	4.0966	0.9337	SMART	366	366	2.5142	0.1138	2.5001	0.1091	TSC	10	356	3.0125	0.1305	2.7258	0.1525	
479 PTC	365	366	2.7987	0.1428	2.9651	0.1490	SMART*	10	356	2.5852	0.1253	2.9605	0.1366	TSE	113	253	4.0188	0.1685	4.6199	0.5396	
480 PTDY	366		2.4399	0.0799	3.5741	0.5146	SNC	108	297	4.6379	0.2021	3.5930	0.4001	TUB	258	365	3.0868	0.3035	3.1722	0.5610	
481 PULSE	242	134	3.4105	0.1709	1.9508	0.0736	SMF	12	354	1.8383	0.0775	3.4143	0.1852	TUR	139	227	1.7302	0.0533	1.8259	0.1609	
482 PURA	339	17	4.1100	0.2312	3.7961	0.2056	SMLY	366	366	3.2007	0.1919	3.9067	0.3087	TWIST	137	7	7.9751	0.3000	4.2296	1.132	
484 PUT	366		2.8004	0.1317	2.9667	0.1462	SNACR	366	366	3.1147	0.1551	3.1518	0.1604	TWLV	366	366	3.3779	0.1763	5.3671	0.3219	
485 PX	99	267	3.2760	0.1247	5.6017	0.8010	SNC	216	152	1.5850	0.0421	2.8529	0.1409	TX	366	2	2.0510	0.0753	3.3784	0.1831	
486 PXC	365	1	2.1578	0.0856	2.6602	0.1186	SND	364	366	2.2981	0.0952	3.2647	0.1669	TCZ	87	279	2.9313	0.3586	4.2009	0.1744	
487 PY	278	90	3.8298	0.1546	3.0581	0.3696	SNGLS	366	262	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
488 PXL	155	211	2.1188	0.0797	2.6972	0.1306	SNIP	93	273	1.8286	0.0625	1.8086	0.0587	U	87	279	2.9313	0.3586	4.2009	0.1744	
489 PYC	366		3.2135	0.1565	3.3843	0.1645	SNL	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
490 PYN	366		2.2861	0.1637	3.0133	0.1539	NOV	263	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
491 Q2C	301	65	3.0453	0.1472	2.2713	0.0967	SNRG	83	283	4.6132	0.1983	2.1485	0.1070	UBQ	366	366	4.0707	0.2182	3.5756	0.1987	
492 QASH	366		2.2940	0.1162	3.3927	0.1775	SOAR	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
493 QAU	123	243	3.3899	0.1362	3.9197	0.3834	SOAR	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
494 QBC	318	48	2.9642	0.1724	3.5449	0.1851	SOCC	321	321	4.6776	0.2763	3.9327	0.2186	UETL	6	360	3.4979	0.1841	3.9123	0.2159	
495 QBK	366		2.7195	0.1665	3.3447	0.1423	SOI	326	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
496 QBT	315	51	3.0637	0.1513	2.5412	0.1149	SOJ	66	300	1.8238	0.0660	2.1472	0.0792	US	72	294	5.1535	0.2416	3.2355	0.321	
497 QCN	174	192	1.7655	0.0529	1.7199	0.0575	SOLE	366	366	17.3848	1.1615	3.3329	0.1805	UKG	366	366	2.9597	0.1180	3.6466	0.1996	
498 QORA	366						SOON	303	63	2.2370	0.0905	3.9716	0.2221	UMO	324	42	3.6650	0.1992	7.473	0.4455	
499 QRK	372	44	3.3544	0.1731	3.6934	0.2002	SOON	303	63	2.2370	0.0905	3.9716	0.2221	UMO	324	42	3.6650	0.1992	7.473	0.4455	
500 RGL	366		2.2417	0.0908	3.4120	0.1803	SOUL	232	134	4.9502	0.2408	2.8682	0.1187	UNAT	366	366	17.7171	1.821	3.3645	0.1835	
501 RSLV	366		17.1052	1.1388	3.2033	0.1771	SPA	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
502 QSP	366		3.1295	0.1570	4.4477	0.2556	SPA	366	366	17.3513	1.1562	3.2644	0.1758	UNC	104	262	3.0509	0.1552	2.4484	0.1051	
503 QTL	315	51	2.8181	0.1346	2.4025	0.1034	SPACE	130	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
504 QTM	366		3.4711	0.1802	4.3389	0.2503	SPANK	340	26	3.0822	0.1507	2.6006	0.1210	UNI	153	231	3.7294	0.2945	8.2327	0.3013	
505 QTZ	32	366	10.7377	0.5168	1.8621	0.2750	SPC	2	364	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
506 QNARK	366		2.4509	0.1207	3.4323	0.1818	SPR	7	359	3.2752	0.1705	4.1288	0.2282	UNIT	331	35	3.2554	0.1505	2.9883	0.1031	
507 R	366		2.9954	0.1535	3.9822	0.2125	SPF	366	366	3.0527	0.1466	3.1261	0.1634	UNAT	98	268	3.2558	0.1261	1.5904	0.0909	
508 RAC	366		3.5895	0.1864	3.6659	0.2027	SPHR	366	366	4.9312	0.1718	4.4566	0.2681	UNO	70	266	3.2588	0.1597	3.4478	0.0922	
509 RADI	50	250	1.4708	0.1366	3.4883	0.4561	SUKTR	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
510 RADS	366		3.2572	0.1673	3.4403	0.1799	SPM	132	34	1.8208	0.0591	1.8898	0.0677	UNO	36	330	1.7410	0.0997	1.6324	0.1440	
511 RAIN	366						SPORT	174	192	1.7275	0.0522	4.2797	0.2501	UNO	255	111	2.8173	0.1147	3.2224	0.2072	
512 RATIO	301	65	3.0470	0.4326	3.6127	0.5842	SPTS	366	366	3.1889	0.1645	3.8824	0.2297	UBQ	237	129	3.1169	0.1504	2.6989	0.0961	
513 RBIES	288	78	1.9491	0.0690	2.8799	0.1															

SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N	SYM	TradDays	nonTradDays	AlphaP	Sd P	AlphaN	Sd N
601 XBC	366		3.6332	0.1941	2.2170	0.0902	XLC	342	24	4.0241	0.2248	3.7091	0.1992	XUP	55	311	2.6179	0.0872	2.1313	0.2412
602 XBL	340	26	3.0821	0.1548	4.6282	0.2667	XLM	366		4.0871	0.2314	6.3271	0.3885	XVC	215	151	2.2676	0.0847	2.0678	0.0896
603 XBOT	7	366	17.4831	1.1655	3.2650	0.1758	XMCC	365	1	3.3451	0.1748	3.1728	0.1593	XVE	123	243	5.2947	0.2382	6.1101	0.7981
604 XBP	80	286	1.4651	0.0255	2.4618	0.2545	XMG	99	267	4.1682	0.1744	9.1032	1.3505	XVG	366		3.4347	0.1748	4.4130	0.2602
605 XBS	59	307	2.0480	0.0561	1.8993	0.2181	XMO	34	332	1.3356	0.1938	2.5994	0.0839	XVP	310	56	2.0702	0.0793	2.0579	0.0780
606 XBSYS	131	235	2.7362	0.0977	2.8963	0.2682	XMR	366		4.2595	0.2334	4.3851	0.2589	XWC	366		2.4146	0.1043	3.2337	0.1656
607 XBY	212	154	2.0259	0.0744	3.5291	0.1906	XMY	366		3.2761	0.1669	4.8293	0.2854	XWT		366				
608 XC		366	3.1466	0.1514	3.3301	0.1814	XNA		366					XXX	366	366	3.2186	0.1569	3.3820	0.1849
609 XCE	135	231	1.9823	0.0716	2.2035	0.0902	XNC	13	353					XZC	366		4.2506	0.2371	3.1631	0.1621
610 XCI		366	3.2130	0.1561	3.3179	0.1805	XNG	82	284	1.8742	0.0615	1.7969	0.0622	YAC	120	246	2.7564	0.0973	4.5792	0.5659
611 XCN	187	179	2.5383	0.1099	3.6063	0.1999	XNN	170	196	1.8664	0.0605	1.6925	0.0546	YAY	100	266	3.9735	0.1615	3.8520	0.5489
612 XCO	7	359					XNX	116	250	3.4310	0.4109	6.0059	0.2752	YBC		363				
613 XCP	366		3.1075	0.1575	3.1107	0.1543	XP	283	83	1.7033	0.0505	1.7875	0.0600	YES	328	38	4.0707	0.2124	3.5891	0.2066
614 XCPPO	133	233	2.3355	0.0919	3.0914	0.1680	XPB	366	17	5686	1.1716	3.3802	0.1847	YMC	64	302	1.5903	0.0319	1.7279	0.1486
615 XCR	366	366	17.2732	1.1507	3.3819	0.1849	XPD	209	157	3.6611	0.2029	3.5781	0.1851	YOC	328	38	4.0971	0.2296	3.1495	0.1585
616 XCRE	2	364					XPH	366	366	17.3513	1.1562	3.2646	0.1758	YOVI	352	14	1.3372	0.0257	1.3856	0.0277
617 XCT		366					XPM	366		3.0802	0.1582	4.9957	0.2876	YOYOW	366		4.7367	0.2642	3.7814	0.2159
618 XCTT	3	363					XPO	2	364					ZAP	272	94	2.5863	0.1157	2.5894	0.1191
619 XDB	48	318	4.5513	0.1918	5.6562	0.9709	XPPOKE		366					ZBC	233	133	2.0570	0.0825	2.0933	0.0769
620 XDE2		366					XPRO	2	364					ZCC		366	18.4054	1.2307	3.3659	0.1836
621 XDN	366		4.2829	0.2440	4.3962	0.2497	XPS	183	183	1.7094	0.0513	2.9350	0.1463	ZCG	95	271	1.7321	0.0535	1.7187	0.0537
622 XDP		366	17.5904	1.1731	3.3649	0.1836	XPT	29	337	2.1956	0.0643	3.1541	0.4817	ZCL	366		3.1634	0.1582	3.2894	0.1711
623 XDQ		366	17.1052	1.1388	3.2823	0.1771	XPTX	4	362					ZEC	366		3.6219	0.1917	9.0239	0.5997
624 XEL	366		4.2754	0.2358	3.3357	0.1776	XPV	80	286	6.7310	0.3099	5.1391	0.8449	ZECD	293	73	3.0174	0.1345	4.0897	0.2602
625 XEM	366		4.2576	0.2339	3.9187	0.2226	XQN	98	268	3.7564	0.4727	4.1995	0.1756	ZED	366	366	3.2108	0.1563	3.3740	0.1843
626 XEN		366					XRA	123	243	2.6395	0.0922	1.9418	0.1332	ZET	235	131	3.2676	0.1396	3.7169	0.2690
627 XFC		366					XRE	365	1	2.9381	0.1449	2.5652	0.1145	ZEN	366		3.2078	0.1602	3.1038	0.1586
628 XFT	274	92	3.0738	0.1478	3.4411	0.1878	XRL	115	251	2.7721	0.0982	2.5039	0.2378	ZENI	243	123	6.3276	0.3835	3.3891	0.1816
629 XG		366	17.1788	1.1412	3.2603	0.1760	XRP	366	359	3.0906	0.1537	5.1466	0.3082	ZER	365	1	2.0882	0.1385	3.4781	0.1847
630 XGB	14	352					XSEED		366	3.2270	0.1575	3.3345	0.1812	ZET	164	202	1.7144	0.0422	2.7804	0.2003
631 XGOX	357	9	1.8084	0.0555	1.7997	0.0644	XSI		366	3.2176	0.1568	3.3108	0.1794	ZET2		1				
632 XGR	84	282	2.1665	0.0844	2.2935	0.0978	XSP	7	366	3.7679	0.2075	3.7006	0.1970	ZLQ	126	240	3.0623	0.1153	3.0276	0.2990
633 XHI	228	138	2.9289	0.1201	2.8149	0.1746	XSPEC	366						ZNY	47	319	1.5994	0.0323	1.8481	0.1851
634 XID		366					XST	366		2.9520	0.1476	3.8248	0.2044	ZNY	47	319	1.5994	0.0323	1.8481	0.1851
635 XID*	31	335	1.4470	0.0239	1.5195	0.1299	XSTC		366					ZOJ	273	93	2.1995	0.0877	3.7421	0.2050
636 XIN	362	4	2.1267	0.0880	2.1752	0.0827	XT	114	252	5.3853	0.2448	11.1155	1.5079	ZOOM	42	324	6.2526	0.2808	5.3396	1.0849
637 XIN*	289	77	1.3020	0.0249	1.4099	0.0277	XTC		366	17.3027	1.1528	3.4022	0.1864	ZRC	366		6.7526	0.4379	3.4593	0.1779
638 XIOS	124	242	3.1461	0.1203	3.7150	0.3925	XTD	20	346	2.9752	0.1051	6.1518	1.4289	ZRX	366		4.2963	0.2513	5.1588	0.2886
639 XJO	76	290	3.4072	0.1305	4.1292	0.6137	XTZ	233	133	3.2951	0.1418	2.7729	0.1739	ZSC	366		2.4806	0.1094	3.4605	0.1819
640 XLB		366	3.1836	0.1544	3.2701	0.1762	XUC	366		3.7465	0.1993	3.4375	0.1837	ZSE	274	92	2.2122	0.0864	2.9232	0.1479
641 XBC	366		3.6332	0.1941	2.2170	0.0902	XLC	342	24	4.0241	0.2248	3.7091	0.1992	ZUR	107	259	3.8970	0.1607	4.8615	0.6031
642 XBL	340	26	3.0821	0.1548	4.6282	0.2667	XLM	366		4.0871	0.2314	6.3271	0.3885	ZXT		366				
643 XBOT		366	17.4831	1.1655	3.2650	0.1758	XMCC	365	1	3.3451	0.1748	3.1728	0.1593	ZYD	85	281	3.4086	0.1304	3.8671	0.5734

Table 18 Summary-list of the cryptocurrencies considered on the market analysis along 2018 (part 5 out of 5)